说　明

时代的发展，科学技术的进步，对人的能力的要求越来越高了，这就向教育提出了激励性的挑战。教育要改革、要提高、要注重培养学生接受新知识、解决新问题的能力，这才能使学生面向未来，做时代激流的勇进者。那么如何回答这一挑战呢？这正是当前世界瞩目、各国教育改革中急待解决的大问题。在时代精神的鞭策下，世界上，尤其是美国的知名心理学家纷纷转向这个重大问题的探讨领域。他们把现代的学习理论和认知科学与教育实际相结合，形成了一个颇具生命力的新的心理学分支——教学心理学。教学心理学的宗旨在于了解和促进人类从不知到知，从低水平向高水平的认知转化过程。教学心理学的诞生标志着在心理学的理论研究与教育实际的结合上向前迈进了一步；意味着人们对“掌握知识，提高能力”的本质和内在规律有了深入的探讨和新的理解。教学心理学的研究必将会为教育改革做出应有的贡献。

为了促进我国教学心理学研究工作的开展，我们译出了《教学心理学的进展》(第一卷）一书奉献给读者，愿能帮助读者从中获益。这本书系由美国教学心理学研究的带头人、匹兹堡大学教授罗伯特·格拉塞主编。书中所选文章均是在探讨心理科学与教育实际的结合上做出贡献的优秀代表作。

本书由杨琦、朱法良、孟鸿伟、赵尚武、富安利、李树珍同志翻译，杨琦同志统校了全部译文。在翻译过程中伍棠棣教授给予了

指导，并校阅了有关章节。

张锦帆、刘英敏同志参加了译文的整理工作，谨致感谢。

译　者

1986年11月

目　　录

序

"教学心理学"(instructional psychology)这个词，第一次被用作标题来给一个研究领域命名，最早见之于1969年的《心理学年鉴》。自此以后，"教学心理学"这一研究领域乃得以发展而日臻形成。它的一般特征是：研究知识和认知技能的获得以及如何设计并安排学习条件来发展这种知识和技能。这个研究领域的兴起，标志着心理学与教育的关系发生了意义十分深远的变化。

在20世纪50年代和60年代初期，实验心理学家一般并不热心研究教学问题。一端是心理学的基础研究；另一端是如何把教材训练的技能、能力学到手的研究；在这两者接界的地方却划定了一个"无人区"，几乎没有人练练功去探索它，或者乐意去考察它。在教学机器和程序学习盛行的那些年月里，有些关心行为科学和教育工艺学之间的关系的人，曾担心：这两者的联系原来就松散，如果仓促从事，一昧迎合眼前的实际应用，就会把教育工艺学远远拉开，脱离行为科学研究。甚至在近10年以前，要吸引一些杰出的心理学家去考虑研究一些哪怕与教育沾点边的问题，也是很困难的。

心理学和教育并不总是如此互相分离的。本世纪初叶，在当时许多大人物的著作中，学习心理学和教育心理学还是密切联系在一起的。多年来，实验心理学和教育心理学不时地相互啮合，为我们展望教学心理学这一新领域的未来提供了一个令人瞩目的前景。本书导言中还将叙述教学心理学的简史。

现在，由于行为主义心理学侵入教学工艺学的影响以及当今在研究人类学习和作业的理论上占统治地位的现代认知心理学的兴起，使得情况发生了实质性的变化。这种变化伴随着研究策略上的改变。在旧的研究模式中，常常把实验室控制条件下所取得的结果径直地外推到复杂的学习中去，而不研究它的局限性和有关的分界条件。现在情况不同了，大部分研究不再用老一套的方法，而是根据人的复杂的、实在的作业提出问题和发展理论的。人们越来越体会到：对较为正规的教学情境中人的知识和能力的获得进行研究，这可能是对理论的正确性及其社会效益的一种重要检验。因此，教学心理学正是在认知科学和学习理论迅速发展的基础上形成起来的。

这套新丛书的目的在于：把那些看来在发展为系统阐明教学问题所必需的知识和学说方面有贡献的杰作，汇编成册出版，借以促进教学心理学的发展。这套丛书打算为各个领域致力于有关教学过程专题研究的心理学家，提供一个论坛。教学心理学正处在一个积极发展的阶段，它集思广益，既得各家的孕育，又将受各家的培养而成长。为此，《教学心理学的进展》一书将尽反映现状全貌之责，而不囿于一家之言。这个研究领域的活动将决定教学心理学未来全部内容的界说与范围，所以全书篇目需要折衷兼顾。约稿的主要标准是：撰稿人一直积极从事高质量的研究工作，在探讨心理科学与教学问题的相互作用方面有着重大意义，并能提出见识广、挑战性强的见解。特别鼓励特约撰稿人将他们当前的工作和思考给大家提供一幅详细的图景，把它同现有的其他工作关联起来，并把它用他们认为是令人振奋、意义明确的适当方式描绘出来。

本卷各章都是当前界定教学心理学的意义和范围的代表之作。第一章由格里诺(James Greeno)撰写。这章是研究知识结构

分析的一个范例。他提出了一个中学几何学科中问题解决的模型。这个模型是用计算机模拟的方法，根据现代信息加工理论的框架开发出来的，并且详细表述了学生解几何题时的思维过程。

第二章由布朗(Ann Brown)撰写。这章论述了某些元认知技能的发展。元认知直接影响着少年儿童的学习和问题解决。这些思维技能使人易于通过预测、检验、监听和现实测验等过程主动控制自己的学习，而这些技能正表明人在各种情境中能有效地解决问题。布朗回顾了关于元认知技能的实验研究、训练学习迟缓儿童掌握元认知技能的教学试验以及这些能力的文化相对性。

第三章由凯斯(Robbie Case)撰写。这章讨论了一些在将认知发展原理用于教学时必须涉及的问题。他采取了皮亚杰(Jean Piaget)关于智力发展的一般原理，并把这些原理组成一套使儿童获得学校通常讲授的技能和概念的最佳教学程序。他还叙述了以探索一种新皮亚杰学派智力发展理论为中心课题的研究。这种理论谈论了一些皮亚杰理论尚未回答的问题：如何识别学校所教技能的运算结构，如何评估儿童现有的运算功能水平，如何把儿童现有的运算功能水平提高到要求达到的水平。

第四章由萨佩斯(Patrick Suppes)，麦肯(Elizabeth Macken)和赞诺蒂(Mario Zanotti)撰写。这一章从前些章讲的认知加工的种种方式转到论述适于组织教学的整体模型。他们详尽地叙述整体模型的两个实例，并且概述了如何用整体模型来系统考虑课程和教学的问题。在第一个实例中，他们开发了一个预测学生三门课程(语言技巧、数学和阅读)的计算机辅助教学中取得进步的模型，并且用这种模型来规定用于这三门课程的时间总量。第二个实例研究了在发展中国家里使用无线电作为教学技术的情况。本章主要涉及到用定量模型使分组学习的教学组织最优化的问题。

第五章由加涅(Robert Gagnè)和比尔德(Jacob Beard)撰写。

他们考查了现行的评估学习结果的测量基本理论和测量技术。他们从智力技能、言语信息、认知策略、运动技能和态度等各个学习作业类概念的关系上对标准参照测量这个概念进行了详细的评述。他们考虑到的问题有：详细说明评估任务的范围，作业评估的质与量诸维度，测验编制的过程，评估测量适切性的测定，以及测验的选择与设计所需要作出的决策。

这卷书的完成，从一开始就是共同努力的成果，也就是说：早在我与出版者厄尔包姆(Lawrence Erlboum)初次洽谈期间，这个合作就开始了。作者们不遗余力地从百忙中抽出时间把他们若干时日关切从事的研究撰写成文，这是有目共睹的。另外，邦德(Lioyd Bond)，奇(Michelence Chi)，莱斯戈德(Alan Lesgold)，米尔曼(Lay Millman)，佩利格林诺(James Pellegrino)和西格尔(Alexander Siegel)认真细致地审阅了早期的手稿，他们所做的工作虽然不如作者所做的工作那么显要，却是必不可少的，十分辛劳的。

所有的具体工作，包括与作者和出版者磋商的事，都是由我的助手朱厄尔(Joan Jewell)一手经办的。如果没有朱厄尔的协助，这本书就不会跟广大读者见面了。我还要感谢多罗拉(Marlene Daurora)，她打印了最后的修订稿；感谢斯坦顿(Pat Stanton)，他打印了文稿并在协助编辑方面做了大量的工作；感谢罗特曼(Donna Rottman)，她为这卷书绘制了全部的图表。最后还要说明一点，这本书的出版自始至终都得到了匹兹堡大学学习研究和发展中心的资助，在此也表示诚挚的谢意。

罗伯特·格拉塞

匹兹堡大学

导言： 走向教学心理学的历程

罗伯特·格拉塞

如今改弦易辙从事一些对研究教学过程有帮助的分析探究工作的实验心理学家，正与日俱增。这种趋势明显地表现在心理学的许多不同领域里，其中包括有关学习、认知作业、人的发展和社会过程的研究等方面。这股热潮的涌起是有许多理由的：基于解决社会有关问题和单纯学科教学问题的崇高要求；基于现时的心理学理论具有研究更为复杂的认知行为的能力；基于确信侧重任务的研究和侧重学科的研究的相互作用对社会和对心理科学双方都有益的信念。

由于许多心理学家对教育现象发生了兴趣，因面基础研究和应用研究之间的界限就越来越不那么泾渭分明了。考察一下最近以来关于人的知识和智力技能的性质和发展这方面的研究，就可以清楚地看到，有关教育各个领域的研究，包括构成智力和能力倾向的基础的各个心理过程的研究，可能既是高度基础理论性质的研究，同时又是指向实践理解的研究。目前这两方面都有着许多令人鼓舞的研究：面向基础理论的研究和面向教育实际问题的研究。这种研究既有助于实际应用，又有助于从理解和改进教学实践的尝试中增长基础知识；并且同时又促进科学理论的发展。安德森〔John Anderson(1976)〕令人信服地阐明了这种想法。他写道：

> 我对维护严密的理论假说，几乎不感兴趣；但我对发展那种能解释重要的经验现象的理论，却怀有更大的热情。所谓能解释"重要经验现象"的理论，我是指：能说明现实世界中出现的问题和实验室中发生的现象的那种理论。现实世界中的问题……应包括如何提高人们学习语言和运用语言的能力、学习课文和记忆课文的能力、推理的能力和解决问题的能力等方面。这反映了我的信条：认知理论的最终裁决者，将是它实际应用的效益。因此，我正打算把我们对所谓理解人的智力的性质这方面的含义所作的解释，更改一下。过去我曾一度认为：理解人的智力的性质这只能是识别出构成认知行为的基础的一些结构和过程。既然是不可能的，我如今就建议，我们所说的"理解人的智力的性质"，这意思就是指要掌握一种使我们能增进人的智力的理论(pp·15—16)。

要发展这种理论就需要选用另一种研究心理学的方法。目前大多数研究方法的目标是按照事实对学习和作业的各种现象进行描述。而我们要考虑的另一种研究方法的目标是探求那种有助于制定获取知识和技能最佳方式的理论和实验。确实，人们越来越关心我们取得有关学习和认知知识的合乎规范、实际应用方面。〔如，阿特金森(R.C.Atkinson)和保尔斯顿(J.A.Paulston)，1972；布鲁纳(J.S.Bruner)，1964；加涅，1977；格拉塞，1976；格罗因和阿特金森，1966〕。这种实际应用性的任务与纯粹描述性的科学是不同的，但是为了给学习和作业的最佳化提供理论基础，仍然需要对构成学习和作业的基础的一些心理过程，作出详细的描述和假设。

I. 现　状

心理学与教育问题的结合目前正在取得进展。这种结合正是

建立一门以学习和认知的研究和理论为基础的教学心理学所需要的。体现这种相互结合的一些特定研究方面包括有：对教材内容进行心理学的任务分析；幼儿教育与发展心理学；从认知过程方面来解释智力和能力倾向；课文的学习和授课的理解；学习结果的评估以及行为的矫正。现在仅就这几个研究方面作一简要的叙述，也就是对正在形成的教学心理学研究领域作一初步的介绍。

教材内容的任务分析

心理学家越来越注重研究比在实验室里所经常研究的相对说来更为复杂的人类行为。他们正在用现代认知理论的基本思想来理解象领会教材内容、识乐谱、证明几何定理和解物理题等这一些与教育实践有关的知识和技能的胜任能力的构成。各种能详细说明胜任能力作业性质的技术，正源源不断地开发出来，这些技术既要能开拓理论概念，取得实验数据，又要能保持现实生活情境的真实性。在这种技术开发的工作中，任务分析起着极为重要的作用；而各种教学任务正借助于模拟技术、信息网络分析、认知过程模型和学习层次来加以研究。对某一特定教材范围内的学习者所习得的全部内容要进行分析性的描述。一个熟练的读者与一个不熟练的读者，一个擅长数学的学生和一个不太擅长数学的学生，一个象棋名手和一个象棋初学者，他们之间的差别是什么？研究者们正是运用任务分析的方法从学习所需基本心理过程（其中包括注意、知觉、记忆和语言）的认知特点入手来描述这些差别的；并且通过任务分析找出教学的结构单元和顺序，使知识和技能的获得更加容易。因为一个人的能力是随着时间的推移而不断变化着的，所以任务分析必须考虑在不同的发展阶段上可用的过程。

幼儿教育和发展心理学

心理学家越来越多地卷入幼儿的教育干预和补偿教育计划的研究中。这种关心教育的动向与人们越来越重视智力发展、重视如何通过教学提高智力的社会气氛是一致的。发展心理学家致力于各种不同的研究并开展了多方面的探讨，这正反映出心理学理论上的不同观点，心理学家和教育工作者的不同兴趣以及教育目标的多样化。目前理论上和实践上压倒一切的问题仍然是智力发展的可塑性、文化影响的实质和发展的关键期，教育干预的时间选择等等。这些关键问题对于以教学或者采取某种周密形式的教育干预来改变或加速受生物学因素和文化因素影响的智力发展进程的一般效果和特殊效果，关系极大。

目前，幼儿教育新方案扩增的势头已经消失，而是致力于设计出一些评估现有幼儿教育方案效果的新方法，并了解有关的教学过程。皮亚杰学派编制的一般任务潜能测验已不再流行，而对于能解释发展变化的不同水平胜任能力之间发生的认知转换，则进行了比较详细的分析和实验研究。对于按认知发展原则设计教学环境的重大意义，正给予更为精确的考虑。

智力、能力倾向和认知过程

智力测验的应用和理论基础问题，个别差异与教学变量之间的相互作用的研究，智力发展过程方面的研究，以及认知作业的跨文化研究，诸如此类的众多因素都正在迫使着个别差异心理测量学传统上各自分离的领域与实验心理学之间发生了破裂。这些因素使得我们有必要也有可能用与前不同的方法来了解智力和能力倾向；并且改革过去那种对个别差异进行教育评估和考查的方式方法。一种对今后的进展可能富有成效的方法是用认知理论中过

程结构的思想去形成个别差异诸变量的概念。目前沿着这些路线进行的研究正加速发展着。正在研究用来测量和界定智力的测验任务作业所包含的认知过程的性质。我们一旦识别出各种认知策略和元认知过程，就要尝试去评估认知策略和元认知是如何受教学的影响的。

授课的理解

关于整段有意义的散文材料的研究，特别是用于教科书和日常书面通讯的材料的研究是大量增加了。这与传统言语学习实验中那些人工材料或者早期心理语言学研究特别设计的句子和小段落，形成了鲜明的对比。目前正在研究课文的组织结构，因为它影响着学习和保持。关于课文组织和意义理解的研究已基本上取代了经典的言语学习研究，为了满足这种新的研究的需要发展了各种相应的研究技术。这些新的研究正在开始直接说明人们是如何学习课文这个问题的。研究人员正在探索阅读者究竟是怎样把事实组织起来的，怎样根据书面信息进行推理并得出结论的，以及怎样更好地教学生掌握加工课文的技能。

目前继续研究初学阅读的过程，例如字母和词的译码、音素分析、造句规则。与此同时，还对关于理解课中高层次语义加工过程，其中包括认知图式的影响、过去知识的内化表象对理解的影响，进行了越来越多的研究。这种对阅读理解的性质所进行的深入细致任务分析以及对学习书面材料的日益重视，形成了一股潮流，这股潮流在探究知识、认知和教学之间的关系方面，起着楷模作用。

学习结果和教育过程的评价

心理测验的性质及其在心理测量学上的基本规则与方法，正

经受着检验。至于成就测量，企图脱离开传统的能力倾向测验而转向更加注意识别出各种胜任作业水平特点的测量技术。为了设计出能用作业标准而不是用相对常模来解释的各种测验，乃开发了许多测量模型。总的目标是把结果的评价与教学的进行紧密结合起来，以便把不断获取的有关学习者获得能力的最新信息提供出来，作为制定教育决策之用。

一个比较活跃的研究领域是详细分析学校学习的条件，更加老练地研究课堂教学过程。过去设计用来评价课程改革的一些研究，首先是想，通过把学生输入的性质与学生输出的质量联系起来的方式来描述学校学习；其次是对介乎输入和输出之间的加工过程只做了很一般的描述。根据某种课堂教学模型就很难获得关于有效的和不大有效的课堂教学过程的详细信息。目前正开发出许多模型试图用学生的初始能力、课堂教学过程变量以及二者之间的相互作用来解释成绩测量中的方差。特别值得注意的是现在有可能把课堂教学过程诸维数的研究与有关教材训练能力获得的基础理论联系起来。每一次努力的结果都将能强化或激起另一些新的发现。

行为矫正

在行为矫正中应用操作理论和系统强化列联，最早见之于临床心理学。这方面在教育上的应用目前也正开始。为教师编写的行为矫正指南到处可见。行为矫正广泛地用于初等学校，而高等教育正以个别化教学系统的形式采用了操作技术。这些应用的成败直接影响着今后的研究。课堂行为矫正的研究以前只是着眼于减少破坏课堂纪律的行为和提高对课业的注意力，而现在则延伸到着重研究作业的智力方面。对行为矫正的认知方面以示范、替代过程和自我强化过程，以及自我调节作业等一些研究为代表，

这些都日益引起人们的兴趣和关注。

Ⅱ. 历史的展望

上面所提到的几个方面大致说明了教学心理学这个正在兴起的研究领域现状的特点。现在简略地回顾一下实验心理学和教育之间的关系史，这将有助于我们正确地了解和认识这个新的研究领域。

本世纪初，许多关心科学心理学发展的著名心理学家认为：在心理科学和心理科学应用于教育、社会实践之间应该保持一种积极主动的交往。卡特尔(Jomes Cattlle)、杜威(John Dewey)、霍尔 (G.Stanley Hall)、詹姆士(William James)、桑代克(Edward Thorndike)和华生(John B.Watson)都这样或那样地指出，探求关于学习、认知和人类发展的知识，从根本上说来，应该同为教育环境的设计建立学科基础关联起来。在早期心理学思潮占统治地位的众多学说中，联想论和机能心理学因其注重实际应用而颇具魅力。就其实质来说，这些理论都承认：能够改造环境以影响人的生活条件，从而促进儿童的发展，增进儿童的心理保健，提高儿童的教育〔博林(Boring)，1950；鲁宾逊(Robinson)，1932〕。

桑代克主要致力学习心理学的实验室研究。他把S-R联结理论应用到各科教材的教学心理学中去，并且他还与武德沃斯(Woodworth)共同研究训练迁移和形式训练原理，这些研究作为一种范例，发生了很大的影响。他的研究以极为直接的方式进行。他认为有些学习规律，如S-R联结律、效果律，和有关迁移特殊性的发现都能直接用来设计教学和教材。桑代克掌握着经实验室检验过的原理和学说，这些原理和学说就能直接用来设计学习条件和课堂教学实践。

杜威反对著名的联想主义倾向。他不只在对人类行为本质的认识上，而且在对心理学与其应用之间关系的看法上，都与桑代克形成鲜明的对照。杜威设想在科学理论和实际应用之间有一个特殊结构，他称之为"连接科学。"这种连接科学将提供一个思想框架。在这个思想框架中，无论从科学研究中，还是从教育实践中获得的知识都能够相互作用，共同积累和彼此修正。杜威在自己的研究工作中决心致力于把科学见诸实践，他明知要冒长期的风险也在所不辞。至关重要的是，只要对人类行为的基本原理哪怕是有一个非常一般的了解，也应立即在实验学校里加以实施。在杜威看来，科学的研究成果总要为自己在实际中得以应用开辟道路，而学校的即时发展常需要依据通用的、更加直观的原则来进行工作。至于实验室的研究结果和有关理论抽象概念的直接应用，是否必然会导致一门无论对实际工作者还是对科学家都通用的教学心理学这门"连接科学"的发展，这就是一个需要进一步研究的问题。为了发展这种连接结构，就需要把桑代克的实验和理论风格与杜威的累积的思想框架结合起来。他们每个人只说明了问题的一半。

尽管美国心理学早期曾经有过实用主义的倾向，但是在把科学知识变成教育实际，以及同等重要的，把实际变成检验科学理论的研究问题这两方面，美国心理学都有一段摇摆不定的历史。由桑代克、杜威以及哥伦比亚、芝加哥的其他一些心理学家所倡导的心理科学和教育实际之间的这种紧密结合精神只坚持了很短的时间。尔后，教育和心理学就分道扬镳了。这两个领域的彼此脱离是由于各自固执己见的活动所促成的。

这段历史简而言之就是：心理学和教育都迫切需要建立自己的学科，显然就几乎没有需要建立彼此之间的外部联系了。这样说一点也不会过分的。一方面，心理学汲汲于成为一门自然科学，

并且要在"硬"科学中取得一席之地。为了达到这个目的，心理学家纷纷走进实验室，发展了各种实验技术，更多考虑的是为理论研究服务，而不是为现实的教育实际服务。而另一方面，特别是教育心理学和教育测验的心理测量学，发现它们所面临的挑战主要来自实际工作者和诸如师资培训、教学方法、课程编制和供学校测验等问题。教育头等关切的事就是把教育专业建立起来，很少考虑花力气去打好学科的根基。当心理学和教育各自着手完成这些不同的紧迫任务时，心理学家和心理学部门也按其擅长人文科学还是自然科学面加以分类，大多数教育心理学家也按其教育方面的才干走进了不同的教育部门。以实验心理学和教育心理学为代表的两种进取方向由于它们的研究风格不同、组成部分不同而表现出不同的特点。

大致从杜威和桑代克的直接影响之后到第二次世界大战（在第一次世界大战期间有所缓和）这个期间，心理学和教育彼此偏狭而互不通气。尽管如此，教育心理学家还是从现有的学习理论中抽取出一般的教学原则，并用这些原则指导课堂教学实验和个案研究，以便向教师说明这些一般的指导原则对他们的教学是有帮助的。而实验心理学家和理论心理学家很少关心也并不置身于教学实践和教材的开发、发展中，他们也很少关心研究教学实际中学习方面和作业方面提出来的问题——而这些问题对心理学理论和实验室的实验都是会发生影响的。在具有人文科学和自然科学特点的"纯粹"心理学部门那里，教育心理学和对教学的关心，是不受人尊敬的。

由于心理学和教育的互相脱节，出现了一些可笑的怪事。在社会上对教育和训练迫切需要的强烈影响下，测验和心理测量学这个领域中，开发出了一种强有力的心理测验的技术和理论，它受到因素分析的支援，面得不到学习与作业的心理学基本理论的

支持。心理测量学是一门工程技术性很强的学科，但心理测量这种技术在实际应用中所提出的问题一般在实验心理学的工作和有关的理论中却得不到应有的反应。相反，关于学习、记忆、问题解决和思维方面的研究，主要是一些富于理论性的和描述性的工作，却极少去为实际应用提供必不可少的思想框架。实际应用可能产生的理论问题和成果也没能向实验室理论和研究结果提出挑战。

第二次世界大战以后和20世纪50年代期间，在心理学和教学工艺学关系的恢复上曾经作过两次大的尝试。第一次是军队因关心训练问题而发起了一项规模庞大的研究。无论是当时知名的，还是后来才成名的心理学家都投入了这一研究。无论是在美国还是在欧洲，他们都用各种不同的观点和方法，其中包括用来分析熟练作业的各种技术去研究教学问题〔格拉塞，1964，1965；梅尔顿(Melton)，1957〕。第二次是斯金纳(Skinner)把操作心理学引入教育领域。在20世纪50年代后期和60年代，教学机器和程序教学引起人们的极大兴趣，这是有文可查的。

对于这两次尝试，我们来看看两位最重要的心理学家所做的说明，是很有意思的。梅尔顿是美国空军人员选拔和训练中心的指导者。这个中心雇用了一大批和平时期集结的心理学家。九年以后这个中心因缺乏持续的资助而宣告解散时(1960)，梅尔顿说了如下一段话：

> 我们的失败证实了心理学家们的告诫：以现代心理科学能够支持的这种技术，还是很原始的——仅仅是一种初步的估计，而基础心理科学的某种关键领域的真正支持，应该一开始就与注重技术应用同时并举，如果技术是可靠的和有效的话。……现在我们亲眼看到人们在精心地进行技术应用……，但是支撑技术的科学基础处于饥饿状态之中。

象这样重心理技术学而轻其科学基础的例子，这不是唯一的。我在这里再一次想到了心理测验、进步教育、指导、直接教育。我切望今后的10年这种历史不再重演〔p，633〕。

在1965年，斯金纳也觉得应对操作心理学在教育上的应用表示一点关心：

心理学的一个特殊分支，即所谓行为的实验分析，已经诞生了，如果它不是一种艺术，也至少是一种教学技术，人们根据它确实能"推想出教学的大纲、日程和方法"。大家已从教学机器和程序教学这两项成果中了解了这种技术。它们的兴起犹如昙花一现……。遗憾的是，很多这种技术已经脱离了它的基础科学。

教学机器被普遍误解了……，程序教学也被普遍误解了。对于根据行为实验分析所产生的第一批程序只是进行了某种表面的模仿〔pp.427—428〕。

梅尔顿和斯金纳的关心为我们今后的工作提供了一个值得认真考虑的教训，如果过急地使新的应用脱离开它的理论基础，那么这种应用就会变得极为肤浅而宣告夭折，同时理论也得不到提高。通过适当的环节使应用与理论保持联系，这将有助于互相修正，互相补充，这样，无论是应用的失误和局限，还是理论的失误和局限，都能被及时发现，并得以纠正。

在20世纪50年代后期，由居领导地位的科学家和数学家主持了课程设置的改革。布鲁纳(1960)撰写的教育过程一书已大量发行，这是由持不同理论观点的各学科的专家、专业教育工作者和心理学家共同参加的一个会议的总结。布鲁纳写道："看来好象很奇怪，把心理学家和居领导地位的科学家聚在一起来讨论他们各

自学科的教学问题，这还是第一次〔p·ix〕。"20世纪60年代的课程改革从各种不同的观点出发进行了广泛的讨论和争论。根据教学心理学的观点看来，从这些致力于课程设计的改革中还不能产生出一门强有力的有关教材掌握的心理学。但是，学科教学与某种有影响的心理学理论结合成一个整体，这正是加涅〔加涅1970；加涅、梅厄(J.R.Mayor)，加斯坦斯(Gasstens) 和帕拉迪西 (Paradise)，1962〕提出的学习层次论的思想。美国科学发展协会已经用这种理论的思想为小学制定自然科学课程。

为了把教育和心理学结合为一个整体的上述种种尝试虽屡遭挫折，但仍然受到社会和科学的时代精神的激励。社会迫切要求教育制度的改善，教育工作者要求研究和发展教育，而许多心理学家也直接或间接地发现如此结合是符合他们的意愿的，也可以对他们的研究工作进行合理的检验。正是在这种情况下，教学心理学这个新的研究领域得以形成。1964年，希尔加德(E.R.Hilgard)任全国教育研究会年鉴的学与教的理论卷的特约编辑。这卷书的全部章节都是由一些杰出的心理学家撰写的，其中布鲁纳论述了教学理论的性质，并在描述性的学习理论和规定性的教学理论之间做了明确的区分。

希尔加德于1948年出版的引人注目的经典教科书《学习理论》，1966年出版了第三版〔希尔加德和鲍尔(G.H.Bower)，1966〕，其中补充了关于《学习和教学技术》一章，足见这个领域的活跃。在较近的1975年的版本中(希尔加德和鲍尔)这一章改为《教学理论》，在修订中，除了旁的内容以外，还补充了加涅的层次理论，布鲁纳的认知发展理论，阿特金森的最佳化学习的决策理论分析，卡罗尔(Carroll)的学校学习模型，以及斯金纳的程序学习。

加涅和罗沃(Rohwer) 在1969年的《心理学年鉴》上发表的一篇题为《教学心理学》的详述，才称得起是这个领域的第一篇论文。

这篇评述反映了当时学习领域的状况，其中谈到了这样一些课题，如注意和定势、预先训练和迁移、学习的刺激特点、反应的条件、学习辨别、概念和规则中的提示和指导、反馈效果、影响言语学习中保持的变量的研究，还谈到了关于阅读和语言、守恒任务训练等几项研究。格拉塞和雷斯尼克(L.B.Resnick)在1972年的《心理学年鉴》上发表了教学心理学的第二篇评述，他们确定了教学心理学的研究范围，正如本文开头所说的那些，这就给教学心理学的研究领域继续进行了界定。麦基琴尼(W.J.Mckeachine)、威特罗克(M.C.Wittrock)和拉姆斯戴恩(A.A.Lumsdaine)分别于1974和1977年撰文论述教学心理学，并提出了其它一些方面的研究。最新的专题论丛继续反映了为建立一门强有力的教学心理学所做出的探索〔如，安德森、斯皮罗(Spiro)和蒙塔古(Montague)，1977，克拉尔(Klahr)，1976〕。

Ⅲ．教学心理学的进展

上述情况表明，着手汇编、出版一套反映教学心理学进展的丛书，正是时候了。这套丛书的编纂正是在发展变化中开始的。目前，现代认知心理学已成为心理科学中占统治地位的理论力量。而前些年则不同，那时是行为主义的学习理论占优势，行为主义心理学的影响广为流传，并渗透到许多方面，认为行为的学习和重新学习是一种极为重要的现象，特别是在各种治疗情境、收容环境、特殊教育，以及各种水平的“个别教学”中，尤显重要。比起行为主义心理学来，认知心理学在它的研究成果和技术应用于人们实际工作的广度和深度上，还只不过是一只刚会飞的小鸟。

谈谈这种形势是很有意义的。因为注重应用问题对现代认知理论的发展是具有重大影响的。应第二次世界大战中的急需而发

展起来的关于复杂的人—机系统所需技能的心理学研究，促进了认知研究与根据信息加工系统、概率论、计算机模拟、人工智能思想所开发的人类作业模型之间的紧密联系。在这种意义上说，今日认知心理学的主要理论思想，在某种程度上，是出自于对复杂的人类行为的实际关心——这些行为的复杂水平远远超过了那些行为主义理论赖以产生和发展的行为。现时，多方面的实际应用也正在改变行为主义的理论，而认知理论也正是在研究诸如问题解决、语言发展、思维、理解、记忆、表象、智力和授课理解等人类作业复杂领域的过程中，奋力组织自己的理论体系。或许可以从中得到一个重要启示，那就是不仅基础科学研究和理论对应用是有用的，而且实际应用对心理学知识和理论的形成也有重大的影响。为了互相促进，共同提高，似乎需要通过某种媒介和桥梁使理论和应用之间不断地接触。可以预料教学心理学这个领域中的每一进展都会对此做出有益的贡献。

杨琦 译 伍棠棣 校

一、一个问题解决的研究

詹姆斯G.格里诺

I.引　　言

教学，特别是数学和自然科学教学的一个主要目标，在于提高学生解决问题的技能。在数学课上，学生为了学习教材所进行的大多数活动中就包含着问题解决，而一些问题又是被用来检验每个学生掌握教材的程度的。起码说，一门课程的教学目标必须是使学生获得解决本门课程范围内的问题的专业知识。同时还要有更加雄心勃勃的目标，那就是要提高学生的问题解决和推理的一般技能。当然，如果学生没有获得解答习题和测验题所要求的特殊知识，那么他们就没学好这门课程。

学生要解决某一方面的问题必须知道些什么呢？其一是必须掌握教科书和其它有关教材所提供的知识。这些学识是必要的，但仅有这些还是不够的。学生除了必须掌握教科书中所给的概念和定理之外，还必须知道怎样用这些概念和定理去解决问题，这种关于"怎样做"的知识就是属于"技能"这个一般范畴的。但是直到最近，关于问题解决所需要的各种技能还知道得很少。

本章的主要内容是关于中学几何问题解决研究结果的一个评述。这项研究是用流行的信息加工理论框架进行的，并对学生解几何题时所用的过程进行了详尽的表述。这种表述采取了所谓解几何题的Perdix 计算机程序的形式。Perdix所表示的过程与学生

解同种问题所用的过程一样，那么 Perdix 就是学生解几何题所需知识的一个候选模型。这项研究还没有做完，我还不能最后肯定Perdix程序就是学生学好几何课所需要的一个完整的知识模型。但是，这个模型确实可以用于大部分几何内容中，现在看来，我可以把这个模型的原则看作是希望学生在学习中学几何课时所获知识性质的合理假设。

把计算机模拟这种方法运用到这个重要学科，部分原因是计算机模拟可以保证模型中假定有的知识成分能足以完成该模型要完成的任务，这对教学应用是非常重要的。如果发现一个程序解某一个问题需要某一种知识，那么就可以理所当然地认为人类问题解决者在解那个问题时也需要那种知识。我们从计算机模拟中所得到的这个结果与教育心理学家所进行的任务分析所得的结果是类似的〔如，兰达(Landa)，1974；雷斯尼克，1976〕。与传统的任务分析一样，计算机模拟也是把任务作业分解成各种重要的成分。一方面要求这些成分能在一个运行着的计算机程序中共同起作用，一方面我们还要提出一个另外的必要条件，那就是这些组成成分确实足以完成这个任务，而且成分之间还要互相协调，以便能统一在一个功能系统中。

这项研究的一个重要目的是使这个问题解决模型可以作为学生思维过程的一个逼真的表述。为了达到这一目的，我们用了六个正在学习几何课的九年级学生大声思维的记录材料来指导这个模型的开发。在八个月中我差不多每周与每个学生交谈一次，时间为15至20分钟，每次都让学生做几道题。给他们出的题目不是课本中的问题，但是相当于他们的课程进度水平的问题。这些记录材料就成为本研究的基本资料。除此而外，还得到了供进一步实验，用来设计任务以弄清问题解决过程的特殊方面的资料。

这里所作的表述比起通常学校学习和课程的讨论来是过于详

细了，但是，这种详细的表述是必要的，这完全是为了清楚地了解学生获得知识的性质和解决问题所用的方法。在为应用提供基础知识的科学中，基础科学的分析确实包括了许多还不能直接用上的细节。例如与电机工程有关的电学理论；与医疗实践有关的生理学。认知过程的科学分析还处于早期发展阶段，随着研究的深入，我们会找到许多方法来表明这些基本原则而不必要象现在这样烦琐。但是如果使理论内容能充分表达出来，目前我们有必要对各种理论系统做比较详细的描述。

本章的第二部分概要地叙述了关于问题解决心理学的背景材料。当然这不是一个完整的回顾（如需有稍完整的叙述请参阅格里诺，1977a）；但是却提到了我认为是一些引人注目的最新发展以及与后边的论述有关的材料。第三部分概述了 Perdix 模型，包括 Perdix 解一个简单问题的梗概，以便对用模型来表示解题过程的方法提供一个总的概念。第四、五、六谈到了这项研究的主要结果，叙述了 Perdix 解几何题所需要的主要知识。第四部分谈到了知觉模式识别的知识。第五部分谈到了进行推理所需要的定理知识的表述。第六部分谈到了策略知识：问题的目标和子目标以及制订解题计划所用的知识的表述。第七部分谈到了研究结果对实际教学的意义。第八部分谈到了研究结果对问题解决的一般理论及意义。

Ⅱ．一般背景

近年来，认知心理学家对许多认知任务所涉及到的心理过程进行了详细的分析和描述。前几年，心理分析的共同格调是详细说明影响知觉、记忆、学习、问题解决以及研究者正研究的任何心理过程的效果的各种变量，现在，研究者们渐渐地不再去分析那

些影响心理过程效果的变量了，而是详细说明他们认为这些过程是如何进行的各种模型，包括各种组成子过程的性质以及把各种子过程组成一个完整的解题过程的方法。

在问题解决研究中所出现的这种更加详尽地说明过程的趋势，不亚于认知心理学的其它领域。继纽厄尔和塞蒙(Newell,Simon 1972)的基本贡献之后，问题解决的心理分析现在已注意到如解释信息、确定目标、选择有效的解题动作等这样一些组成过程的性质和组织的研究上。这与行为主义和联想主义思想在实验心理学中占统治地位时所进行的研究截然不同。行为主义和联想主义所进行的问题解决研究主要是找出有助于解字谜和概念识别的实验条件。还有一些研究是根据行为主义的强化和消退假设来对功能固着进行了概念性分析。总的说来这些分析比起我们现在所开发的详细分析来是太抽象了〔伯恩、埃克斯特兰德、多米诺沃斯基(Bourne、Ekstrand、Dominouski，1971)的综述是具有代表性的〕。

象邓克(Duncker，1945)、柯勒(Kohler，1927)和沃特海默(Wertheimer，1959)这些格式塔学派的心理学家早就致力于对问题解决的过程进行分析。这些研究把问题看作是情境，由于个体不理解情境中的各种成分之间的重要关系，因而就不能正确地解题。他们对问题的突然解决、对明显的顿悟体验特别感兴趣。邓克的功能固着概念敦促我们注意，由于一种情境成分的某种限制性的概念化而没有办法使这种成分与问题的其它成分有机地联系起来，因而也就不能达到理解。

大多数流行的问题解决心理学研究把找出答案看作是搜索的成功。在许多问题情境中都有一个详细说明的初始情境和目标。问题解决者有一套能用来改变情境的操作。要解题就必须找出把初始情境变成目标的操作顺序。因为有几个不同的操作可以运用，所以就有许多不同的顺序可以试试。但是这些顺序中大多数是不

合适的 也就是说是得不到答案的。因为只有几种（或许只有一种）顺序可能达到标目，以所要找出一种成功的解题路径是十分困难的。

在把问题解决分析为在可能的操作顺序的巨大空间中所发生的搜索时，发展出了许多重要概念。一种想法是手段—目的分析，也就是在组织解题时所用的一般探试。

在分析中，问题解决者把眼前的情境与目标情境加以比较以便找出它们之间的差别。所找出的这些差别就成了解题的焦点，问题解决者提出了若干子目标试图缩小这些差别。由一个决定哪一个子目标先达到的优先系统进一步组织了问题解决。在手段—目的分析中，由于知道了哪个操作与哪个问题解决的子目标有关，因此就减少了需要考虑的可能动作数。在称之为一般问题解决者这个问题解决系统的开发过程中也发展了手段—目的分析概念、差别排序的概念和把操作与差别联系起来的概念。〔厄恩斯特(Ernst)和纽厄尔，1969；纽厄尔和塞蒙，1972〕。

还有一类重要概念涉及到制订解题计划的一般方法。一般问题解决者所发展出的一个想法是通过考虑情境和目标的某些主要特点并制订出一个能消除这些主要特点差别的计划，从而找出一个解题的一般轮廓。问题的最终答案必须有待于消除制订计划时所忽略的那些具体特点的差别，但是有一个适当的计划会极大地促进问题的解决(纽厄尔和塞蒙，1972)。制订解题计划常常需要真正了解问题的范围，包括要知道所要解决的主要子问题〔萨塞多蒂(Saceroloti)1975〕，要知道能被开发并照搬使用的各种部分解法〔法尔曼(Fahlman)1974〕，要知道各种过程的缺陷，有时也叫做“错误”，这会使问题解决者既能预见，也能避免节外生枝，从而顺利地解题〔萨斯曼(Sussman)1973〕。

在最近关于问题解决理论的研究中所发展的第三类重要概念

涉及到问题解决者在连续的解题过程中必须存贮在记忆中的那种信息。在大多数问题解决系统中是按照一系列子目标来组织解题的，这就要求对已试过的子目标做个记录。这个记录对于把握住解题方向是很重要的。当为了达到一个近期目的而建立起一个子目标时，那么在这个子目标实现后，问题解决者必须转向下一个较高的子目标。一个子目标实现情况的记录可以为问题解决者提供信息，以免在同一个不能实现的子目标上反来复去地兜圈子。在某些实验中，分析已开始把信息存贮的必要条件与记忆过程的一般模型和通道容量联系起来〔阿特伍德(Atwood)和波尔森(Polson),1976〕。

在问题解决理论方面得到十分猛烈发展的是所谓的问题解决技巧。关于在各种情境中成功的问题解决所遵循的过程已有一大批概念。目前对所谓问题解决语义学还很少注意，包括如何表示问题，以及如何用问题中的各种定义关系来理解问题并找出答案。但是已做出了一些重要的起步。在对一个问题的课义理解过程的分析中，记述结构与个体对任务的表述之间的重要关系已开始显露出来〔海斯(Hayes)和塞蒙，1974；塞蒙和海斯，1976〕。在对某些初等数学问题的实验分析中，证实了找出问题困难条件和过程的重要性(雷斯尼克和格拉塞，1976)。另外，重要的理论分析表明，分析问题情境模式的优良过程能大量降低对记忆存贮的要求〔赖特曼(Reitman)和威尔科克斯(Wilcox),1976；塞蒙，1975〕。这些分析表明用各种巧妙的概念来表示问题情境有很大的优越性，可以大量节省问题解决的加工要求。

Ⅲ. 关于Perdix的概述

书写Perdix所用的形式结构是一个产生式系统，就是把每个

知识成分表示为包含有一个条件和一个动作的产生式。当程序运行时，其运行过程就是一系列的循环。在每次循环中各个产生式里的各个特定的条件，都得到了检验。最后总发现许多产生式中有一个产生式的条件是真的，那么就执行了这个产生式的动作。执行了一个动作就完成了一次循环。在下一次循环中，还要检验各产生式的条件，直到找出其中一个产生式的条件是真的，然后执行那个产生式的动作，就又完成了一次循环，如此等等。产生式系统这种形式对于建立心理学理论来说是很有用的，因为用各种基本产生式很容易分辨出过程的各种成分，而且还会对过程的不同成分相互作用的方式有一个比较明确的详述。安德森(1976)、亨特和波尔特罗克(Hunt、Poltrock，1974)，克拉勒和沃利斯(Klany、Wallace，1976)，纽厄尔(1972、1973)、纽厄尔和塞蒙(1972)，以及塞蒙(1975)都对作为心理过程模型的产生式作了普遍的讨论。

在表1.1中举了一个产生式的简单例子。在一个人想要进入自己的房间这种情境中就应用了这些产生式。在每个产生式中给了条件，用一个箭头把条件与这个产生式的动作分开。假定开始门是关着的而且锁上了，那么检验P_1，P_2，P_3和P_4的条件，结果都不合适，只有P_5这个缺席条件是合适的，因此第一个循环的动作就是找到钥匙并拿到手。在第二次循环中，P_1、P_2和P_3的条件都不合适，只有P_4的条件合适。这时，所执行的产生式的动作是把钥匙插到锁里。在第三次循环中，P_1和P_2的条件不合适，而P_3的条件是合适的，因此就执行了转动钥匙的动作。在第五次循环中，门已打开了，恰与P_1的条件相符，因此所执行的动作是通过门走进房间。

表1.1　一个简单的产生式系统

产生式	条件	动作
P_1	门开着→	进门，关上门。
P_2	门未锁→	转动门把手，开门。
P_3	钥匙插在锁里→	转动钥匙。
P_4	钥匙在手里→	把钥匙插在锁里。
P_5	其它→	找到钥匙，拿在手里。

我们看到这个产生式对某些情况做了适当的解释。例如，每个循环总是以检验门是否已经开着开始。因此，如果一开始门就是开着的，那么这个系统就不需要执行那些不必要的步骤，象找钥匙、开门上的锁等。另一方面，我们故意使这个例子简略和不完整，以便浅显易懂。实际上打开锁和开门所用知识的一个严肃的心理学理论应当包括找钥匙，检验通过转动钥匙是否已成功地打开门以及其它成分这些例行程序的详细表示。

用Perdix 表示几何知识的产生式系统包括若干个检验问题信息模式的条件和生成新的信息和关系的动作。我们打算用 Perdix 来表示基于模式匹配和模式填充过程的问题解决。这个模型用希腊神话中的人物珀迪克斯命名，珀迪克斯是希腊神话中的建筑师、雕刻家代达罗斯的徒弟，他受一副鱼骨头架的启发而发明了锯，受蟹钳的启发而发明了两脚规。这里所使用的模式这个概念是很一般的。有些模式是图形的各种视觉特点的组合，有些模式是一般概念，如全等三角形的概念。有些模式是抽象的一般结构，象在一个证明中各种定理之间的关系模式。

书写Perdix所用的一般程序设计框架是由安德森(1976)开发的ACT系统。用ACT框架书写的产生式使用有标号的网络形式来表示。任何时候，问题情境都被表示为一个网络。节点代表问题

情境的各个成分，网络上的节点互相联结起来的方式就用来表示模式。在几何问题成分中，有些成分是图形各种组合，如点、线段、角、三角形。网络上的另一些节点是一些一般概念，象全等或对顶角这些关系的名称。用与问题情境相当的方式把这些节点连接在一起。例如，把表示一个角的节点与表示这个角的顶点和两边的节点连结起来。如果发现两个角有对顶角的关系，那么表示对顶角关系的节点就与表示这两个角的节点连接起来。

当Perdix开始解题时，有一个表示问题初始情境的网络。这包括图形所提供的信息以及其它已知信息如平行线或平分角这些关系所提供的信息。问题的目标是让问题解决者生成网络中没有呈现的关系。产生式的条件详述了推出新关系所应遵循的模式。在每一个循环过程中，都对与问题目标有关的各种模式进行一系列的检验。当在网络中找到一个合适的模式时，产生式的动作总是要在与所能推论出来的关系相应的网络中加上新的连接。因此，问题解决涉及到通过在网络中增加新的关系材料来改变情境。问题的答案相当于增加问题目标所要求的特殊关系模式。

A.一道例题

现在我用Perdix解平行线问题的草图来解释上述思想。在图1.1的方框(a)中给了一道例题。解这道题就说明了用模式匹配和模式填充解题的一般思想。另外也说明了解几何题所需要的三种主要知识，在IV、V、VI三部分我还要详细叙述这三种知识，现在只做一个简要的叙述。在以下的讨论中，我要用到的问题图形的各个角在图1.1方框(b)中标出了名称。

解这道题所需要的第一种一般知识是学生必须能辨认出有特殊几何关系的角。例如，看到A_1和A_6时，学生应能看出这两个角是对顶角。解题过程的这一部分涉及到视觉模式识别。第二种一

般的必要条件是具有能进行推理的定理知识。例如A_1和A_8相等，因为它们是对顶角。学生正是根据"对顶角相等"这一定理得出这个结论的，凡是学过几何的学生都掌握了许多这样的定理。解题过程的这一部分涉及到进行推理的定理知识。第三种必要条件是能提出目标和子目标并且制定出达到目标的计划知识。这种知识可把问题解决过程组织起来，并且指导着模式识别和推理过程。例如，一个学生必须知道要想得知某个角的度量，就需要知道这个未知角和另一个已知度量的角之间的定量关系，例如相等或互补。问题解决过程的这一部分涉及到提出目标和制定计划，可称之为策略知识。表1.2列举了解图1.1中的那个问题所涉及到的具体策略步骤。

在开始解图1.1中这道题时，问题的初始情境是用一个由各种组合和各种连接构成的庞大网络表示的。在方框(a)中有表示线段a、b、m、n的端点和交点这12个点的节点。还有另外的节点表示这个图形中所给的所有线段和角。表示出由已显现的线段决定的抽象线和从各已知点引出的射线也是很有用的。角的边要被标为射线而不是线段，以避免由线段而产生的某些模糊。(例如，能把角q的水平边看做是只到直线M的线段或从角q的顶点到已知边端点之间的这一线段吗?)这个网络还包括了问题中所给的平行线关系，以及表示为抽象线之间的而不是任何线段之间的关系。在图1.1方框(c)的形式中，说明角P度数的已知信息被表示为代号A_1与数字40之间的MEAS(度量)关系。还看方框(c)，问题的目标也被表示为代号A_{12}和一个代号为?NUM的未知数之间的MEAS关系。

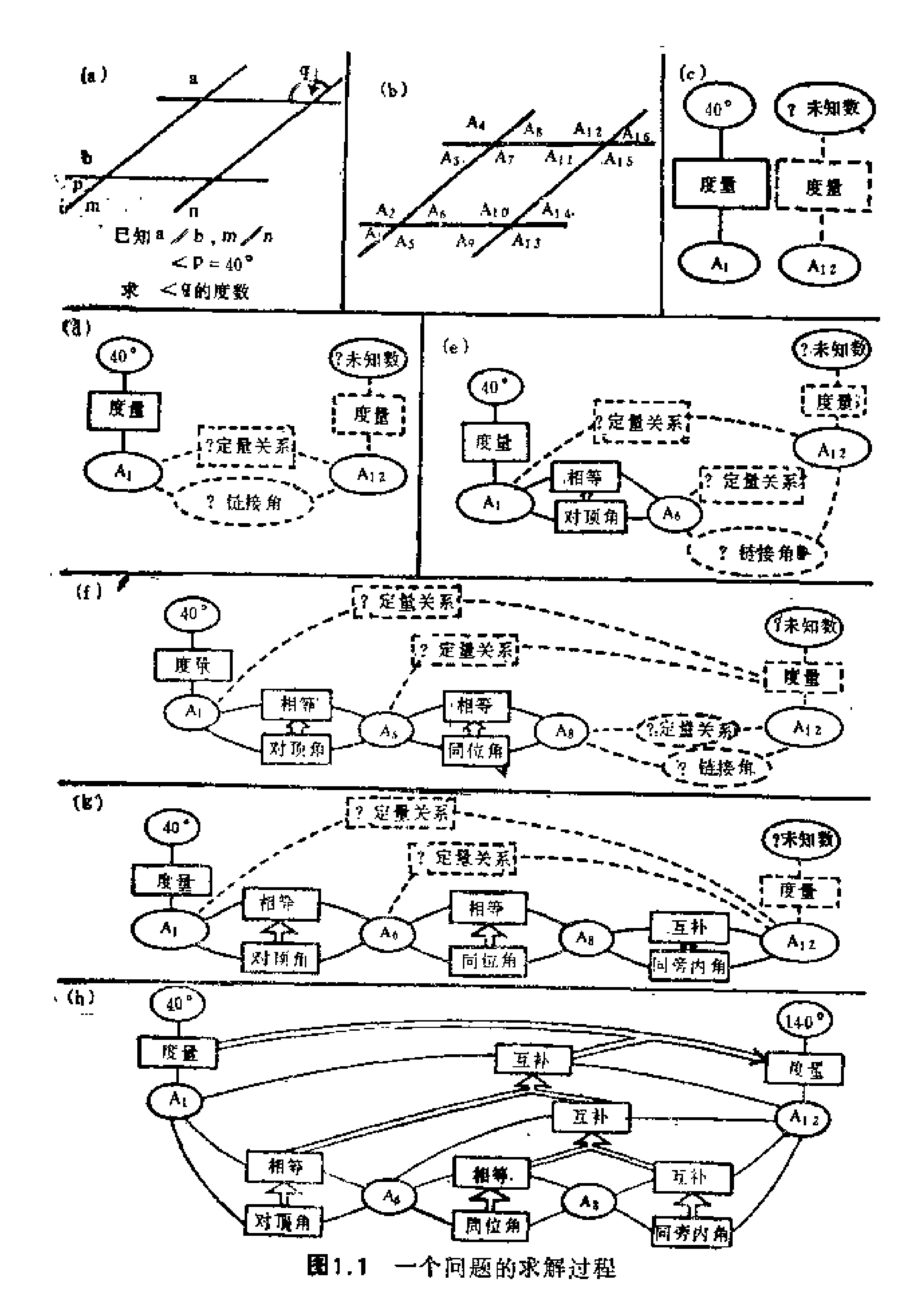


图1.1 一个问题的求解过程

表1.2　　图1.1所表明的提出目标的策略知识

1. 如果当前的目标是求出一个角的度数，那么只要你知道另一个与未知角有某种定量关系的角和度数就可达到目的。如果这样达不到目的，就提出这样一个目标，推断出这个未知角与某个度数已知的角之间的定量关系。
2. 如果当前的目标是推断出两个角之间的定量关系，那么只要知道两个角之间的几何关系或者这两个角都与第三个角有定量关系就可能达到目的了。如果这样达不到目的，就提出这样一个目标：推断出角之间的几何关系。
3. 如果当前的目标是推断出两个角之间的几何关系，那么就要提出这样的目标，使用模式识别的过程分析这两个角的特点。
4. 如果当前的目标是推断出两个角之间的定量关系，并且已经确定未知角和某一个度数已知的角之间没有几何关系，那么就提出这样一个目标，找出一个位置上处于两角之间，并与其中一个角相等的角。做完这一步后，再试着把这个中间角与那个原来目标中的角联系起来，形成一连串的定量关系。

图中用虚线表示的是一个目标，而不是数据结构中的一种关系。在图1.1的所有方框中，用长方形表示的节点代表各种关系，用椭圆形表示的节点代表问题的目标，如角和数量。把一个关系节点与目标连接起来的单线表明用这种关系可把哪些目标连接起来。虚线代表目标。双线箭头表示推断，把一些关系增加到网络中，因为其它关系已经被识别出来了。

图1.1方框(d)表明了解这道题的第一步。在表1.2中，我们看到，第一条策略规则叙述到目标是求出一个角的度数，一个合适的子目标是找出那个角和某个已知角之间的定量关系。因为A_1的度数已知，所以Perdix就提出了这样的子目标，找出A_{12}和A_1之间的某种定量关系（表示为?QREL)。现在，Perdix开始检验有什么几何关系被推出来了(表1.2中的规则2)，结果发现什么关系

也没推出来，那么就分析A_1和A_{12}的特点，看看在它们之间是否能找出象对顶角和同位角等关系(表1.2中的规则3)。这样还不行，于是就使用了第四条规则，Perdix提出了这样的目标，找出一个可以用来展开一连串定量关系的角。在图中一定能找到这样一个角，并用?CHANGLE表示。

当Perdix有了一个找出的目标时，就在要联系起来的两个角之间形成了一条扫描线①，并按接近扫描线的远近检验那些角。而现在这个模型有了一些改进，用扫描机制解释了学生问题解决的某些特点，这是以前的模型所不能解释的。在这道题中A_6和A_{11}都正好在扫描线上，其中A_6与已知角A_1有关，可以推出相等关系(CONG)。结果在图1.1方框(e)中表示出来了。通过知觉模式识别的产生式确定了A_6和A_1具有对顶角的关系，有了这种关系就可以通过一个定理的产生式推断出相等。

当找出一个等于A_1的角时，下一个目标就是分析那个角，看它是否与A_{12}有关。如果这个目标达不到，那么就需要再找出另一个链接角。离扫描路线最近的角是A_{11}，但是通过模式识别的产生式发现A_{11}和A_6没有什么关系。其次是角A_7和A_{10}最接近扫描线，但是Perdix只接受与基角相等的那些角。根据Perdix所进行的计算发现A_3、A_8、A_9和A_{14}这四个角离扫描线的距离相等。四个角中的任何一个都与A_6相等，在这道题中首先考虑的是A_8和A_6是同位角的关系，在方框(f)中表明了所推断出的相等关系。

正如在方框(f)中所表明的，把A_8与A_{12}联系起来的目标就被提出来了。Perdix看到A_8和A_{12}有两个边是平行的，而另两条

---

① 以前所建立的模型中没有扫描这个特点，但在这道题里使用了这个特点(格里诺，1977)。在以前的模型中，一个把角联系起来的链接是通过寻找一个与已知角有特殊关系的角而形成的。

边是共线(CORK)。根据这种特点就可选择一个达到目标的计划，找出所谓PRLREL的定量关系，这是平行线关系的缩写。同旁内角关系表示为INTSAM，由此可推断出补角关系(SUPP)，这些都在方框(g)中表示出来了。

接着，用关系链接推出A_{10}和A_{12}之间的补角关系。根据这种关系和A_1、A_{12}的度数就可计算出来，问题就解决了。在方框(h)中表示出Perdix解这道题的最后关系结构。

B.与学生解题过程的对照

这里所描述的这个研究程序的一个主要目的是开发一个与学生解题方式相类似的解几何题的模型。在图1.1中我叙述了Perdix的解题过程。现在我再谈一下几个学生解这道题的作业，讨论一下学生解题的记录材料和Perdix解题之间的相似之处和不同之处。

表1.3是这组学生中的一名，在学几何课的期间里，按规定时间进行观察所做的一份完整的记录材料。

表1.3　　学生解图1.1问题的谈话记录

*1	S：	嗯，a平行于b，M平行于n，角的度数，噢，角P的度数是40°。(停顿)
	E：	你在找什么？
*2	S：	噢，我可以告诉你答案了。应该……角P与第一个角互补，这个角是……
	E：	好，我来画出你说的这个角，这样我就能跟上你所说的了，角P在这儿。
	S：	对。
*3	E：	这是你刚刚所说的第二个角。
*4	S：	对，它是140°。
	E：	好。

	S：	由于第二个角与第一个角是平行线的同位角。
	E：	对。
*5	S：	第二个角与角g相等。
	E：	嗯。
	S：	因为平行线的外错角是相等的，所以角g将等于140°。
*6	E：	好，现在你看见了任何阶段上的格局了吗？你看一下图，你能告诉我你是怎么把每步连在一起的吗？
*7	S：	噢，实际上我并不是那样做的，我做的第一件事是象这样交叉着看，因为那是你的思路，因为你懂得Pq。
	E：	嗯。
*8	S：	我看到P和对顶角相等，我就接着往下做，我又看到了互补关系。我断定q应当是……我再看看，我断定从P开始往下做很快就能得出答案。
	E：	你找到了一个比较快的解法，只有几步就可以了。很好。

在"*1"处，学生看题。在"*2"处，学生给出了答案并且说明得出答案的理由。此时我打断学生的回答，以便得到一个我以后可以用来记住步骤顺序的略图。在"*3"处，学生首先找出A_5作为关系链接上的第一个角。在"*4"处，学生又找出A_{13}作为链接上的第二个角，并推出度数，说明理由。在"*5"处，学生指出了A_{12}的相等关系，并加以证明，最后推出题目中的未知角A_{12}的度数。在"*6"处，我问学生在求解过程中是否有任何值得注意的子阶段。正如记录材料中所记载的那样，学生又报告了一个首先想到的不同的解法，报告并追溯描述那个原来的解法。在这个解法中，找的第一个链接角是A_8，A_8与A_1是对顶角，第二个链接角是A_{12}，A_{12}与A_6是同位角，第三个链接角是A_{16}，A_{16}与A_{14}是同位角，而A_{12}与A_{16}又是补角。在"*8"处，叙述了这个解题步骤。被试追溯报告中这个首先想到的解答也是Perdix能精确无误地生成的一个解答。但是被试的这个追溯解答与图1.1方框

(f)中所示的步骤有所不同。在首先考虑哪个角作为第二个链接角上Perdix刚好先选A_8作为第二个链接角。而A_8也是与扫描路径等距离的四个角中的一个。而被试选A_{14}作为第二个链接角，这刚好也是Perdix容易首先选择作链接角的一个角。在A_{14}之后，Perdix要么考虑A_{11}，要么考虑A_{18}；如果首先考虑A_{18}，那么Perdix的题解就与这个被试所报告的题解一样了。

表1.4　　学生被试的解题过程

步　骤	被试1	被试2 (最初的报告)	被试2 (追溯解答)
1	(A_1是A_8的对顶角)　(A_1和A_8相等)	(A_1和A_5在同一条直线上) (A_1和A_5成补角) (A_5是140°)	(A_1是A_8的对顶角) (A_1和A_8相等)
2	(A_6是A_3的内错角) (A_8和A_3相等)	(A_5是A_{13}的同位角) (A_5和A_{13}相等) (A_5是140°)	(A_8是A_{14}的同位角) (A_8和A_{14}相等)
3	(A_3是A_8的对顶角) (A_8和A_3相等)	(A_{13}和A_{12}是外错角) (A_{13}和A_{12}相等) (A_{12}是140°)	(A_{14}是A_{18}的同位角) (A_{14}和A_{18}相等)
4	(A_8是A_{12}的同旁内角) (A_8和A_{12}成补角)		(A_{18}和A_{12}在同一条直线上) (A_{18}和A_{12}成补角)
5	(A_1和A_{12}成补角) (A_{12}是140°)		(A_1和A_{12}成补角) (A_{12}是140°)

步骤	被试3	被试4 （最初的解答）	被试4 （修正后的解答）
1	（A_1是A_5的同位角） （A_1和A_5相等） （A_5是40°）	（A_1与A_6是对顶角） （A_1和A_6相等）	（同最初解的第一步）
2	（A_5是A_{11}的同位角） （A_5和A_{11}相等） （A_{11}是40°）	（A_6和A_{11}关系不明） （A_6和A_{11}相等）	（A_6是A_3的内错角） （A_6和A_3相等）
3	（A_{11}和A_{12}在同一条直线上） （A_{11}和A_{12}成补角） （A_{12}是140°）	（A_{11}是A_{16}的对顶角） （A_{11}和A_{16}相等）	（A_3是A_{11}的同位角） （A_3和A_{11}相等）
4		（A_{16}和A_{12}在同一条直线上） （A_{16}和A_{12}成补角）	（同最初解答的第三步）
5		（A_1和A_{12}成补角） （A_{12}是140°）	（同最初解答的第四步） （同最初解答的第五步）

步骤	被试 5	被试 6
1	（A_1是A_8的外错角） （A_1和A_8相等）	（A_1是A_6的对顶角） （A_1和A_6相等） （A_6是40°）
2	（A_8是A_{12}的同旁内角） （A_8和A_{12}成补角）	（A_6是A_3的内错角） （A_6和A_3相等） （A_3是40°）
3	（A_1和A_{12}成补角） （A_{12}是140°）	（A_3是A_8的对顶角） （A_3和A_8相等） （A_8是40°）

4		(A_3和A_4在同一条直线上) (A_3和A_4成补角) (A_4是140°)
5		(A_4是A_7的对顶角) (A_4和A_7相等) (A_7是140°)
6		(A_7是A_{12}的内错角) (A_7和A_{12}相等) (A_{12}是140°)

注：引自格里诺的"问题解决的理解过程。"载N.J.卡斯特兰，D、B、皮索尼和G·R波斯编辑的"认知理论"(第二章)。Hillsdale，N·J·Laureme Erlbaum Associates，1977。

表1.4以简略的形式列出了六个学生被试的解法。表中的每一行都对应于Perdix解题过程中所建立的角之间一种关系。而在被试的解法中这些关系几乎大多提到了。为了保持解法细节上的一致又增加了几种关系，以便与Perdix所能完成的作业进行比较。

在被试1的解法中，如果在找出对顶角A_6之后，首先考虑的是A_8而不是A_3，那么就与Perdix的解法完全一样了。被试2解题过程的原始记录已在表1.3中列出来了。他的最初的解答、追溯的解答以及被试1的解答都包括在Perdix以其现在的形式准确无误模拟的解答中。

被试3的解法与Perdix的解法有所不同。被试3首先找出A_3而不是A_6作为第一个链接角。如果我们编的Perdix程序也首先选择A_3作为第一个链接角，那么所需要的扫描过程就不同了，那就要在被联系的角的顶点之间形成一条路径并进行扫描，选择A_{11}作为第二个链接角，这倒符合Perdix的扫描过程。但是，被试3的记录材料与Perdix的作业稍有不同，通过一个矫正的扫描规则

Perdix的作业可以得到改正。被试3在链接过程中总是先推出每个链接角的度数。而Perdix的规则是接连找出与正考虑的角相等的诸个角，直到有一个与未知角有某种定量关系的角为止。被试3所遵循的策略是，每找出一个与所考虑的角有关系的角都要推算出其度数，直到有一个角与未知角有某种定量关系为止。被试3所使用的策略比Perdix所执行的程序稍复杂一点。因为被试3的策略是每一步都推出链接角的度数，而不是等着找出未知角与最初的已知角之间的定量关系。

被试4解题时犯了一个很有趣的错误，这就从反面支持了Perdix程序解题时执行扫描的假设。在找出A_1的对顶角A_6以后，被试对A_6和A_{11}之间的关系做了错误的判断。事实上，A_6和A_{11}是平行四边形的两个相对的内角，如果有可用的定理，A_{11}当然可以作为链接角。但是，按当时的课程进度还没有学到关于平行四边形的有关定理。这个学生说："有一个定理说这两个角是相等的。"这个学生超越了课程的进度走到前边去了，并且完成了一个解题略图，接着说："除了这个定理之外我还能说出所有的小定理。"被试在解这道题时采用"没学过的定理"作为子问题。并且以标示为修正解答的方式解了题。Perdix也可象被试4那样把A_{11}作为第二个链接角，但是当模式识别系统判断出A_6和A_{11}不能以任何已知的方式联系起来时，就避免了这个错误。被试4可能由于A_6和A_{11}这两个角明显的相等而犯了错误。总之，只要把A_6和A_{11}联系起来就会出现这种情况。被试4所犯的错误似乎支持了这个假设，即扫描过程是寻找一系列链接角过程的一个必要成分。

被试5的解法是Perdix做不到的。他所用的策略显然比Perdix更有效。Perdix是从已知角开始向未知角进发，在这之间建立起一个有关角的链接。在这个求解过程中，某个角如果与Perdix当前所注意的角相等，那么这个角就被包括在这个链接中。被试5的

解答似乎只需要找出一个链接角，把已知角与未知角联系起来。把这种更加有效的策略加入Perdix现有的策略知识中不是很困难的，但是我没有那样做，因为在我所搜集的被试的记录材料中，这种情况出现得不太多。

被试6的解题过程与被试3类似，每步都推出一个度数。寻找链接角的前三步与被试1的顺序一致。但是，在第四步被试6就与被试1及Perdix不同了，被试6没有找出A_8和A_{12}之间的同旁内角关系。因此，被试6就不象Perdix那样只使用相等角的策略来求未知角。但是，如果每一步都要推出一个度数，那么Perdix的策略原则就似乎不可能了。做完第三步以后，被试6说："我要能找到一个角等于q就好了，但是我没有办法找到，因此我只好使用补角。"如果模式识别系统不能找出同旁内角的关系，那么自然应该想到A_8的补角是A_4。但是还不清楚为什么被试6没有看出A_8和A_{12}之间的同位角关系，虽然我们可以猜测先使用A_8在某种程度上干扰了扫描过程。

总而言之，Perdix的解题过程与这些学生的解题过程大致雷同，六个被试中，被试1和被试2首先提出的解法如果以适当的顺序注意角就完全可由Perdix模拟了。被试5的解法使用了一个更简单有效的查找策略，同时寻找出一个即与已知角有关系又与未知角有关系的角。被试3的解法策略上有点变化，使用了不同的扫描过程，每一步都要推出一个度数。被试4的解法犯了一个错误，这个错误似乎出自与Perdix的扫描过程一致的过程，当必要的特点只进行了部分匹配时就承认有某种几何关系。被试6每一步都推出一个度数，显然没有识别出两个角之间的关系。上述这些不符之处的存在确实说明了Perdix不是学生问题解决过程的精确副本，但是，要精确模拟学生关于这个问题的作业似乎只需要做些题目上的变动就可以了，而不需要做结构上的变化。因此，可以

合理地推测出在Perdix中所表示的知识是学生成功地解这类问题
时所获得的知识和策略的一个极好的近似，极其类似于学生被试
正确地解几何题所使用的知识和策略。

Ⅳ、知觉模式识别

在这一部分和下面两部分，我将比较详细地讨论Perdix解几何题所使用的各种知识结构，在这个部分先讨论视觉模式识别的模拟。为了清楚地说明模式识别过程，就必须先详细地叙述Perdix表示信息的方法。

A.图形信息的表示

Perdix把一个图形中的信息表示为一个网络，在网络中用节点来表示点、线段、角和其它几何图形、用链来表示这些几何图形之间的关系。图1.2中的这个例子就给出了插页上那个简单几何图形的网络表示。图1.3标出了图1.2中表示的那个几何图形的节点。节点PNT159、PNT160、PNT161、PNT170分别表示点M、Q、N、P。(表示节点名称的数字除了代表不同的节点外没有其它的意思)。请看图1.2，图中表示每个点的节点用一个链与其标号连接起来。每个点的X和Y坐标也都被存贮起来，但在图形中没有表明，Perdix解题时不使用它们。节点SEG166、SEG167、SEG169和SEG174表示在这个图形中所能看到的线段。每条线段都与它的标号和两个端点连接起来。节点RAY163、RAY164、RAY165、RAY168、RAY172、RAY173表示这个图形中由线段所决定的射线。

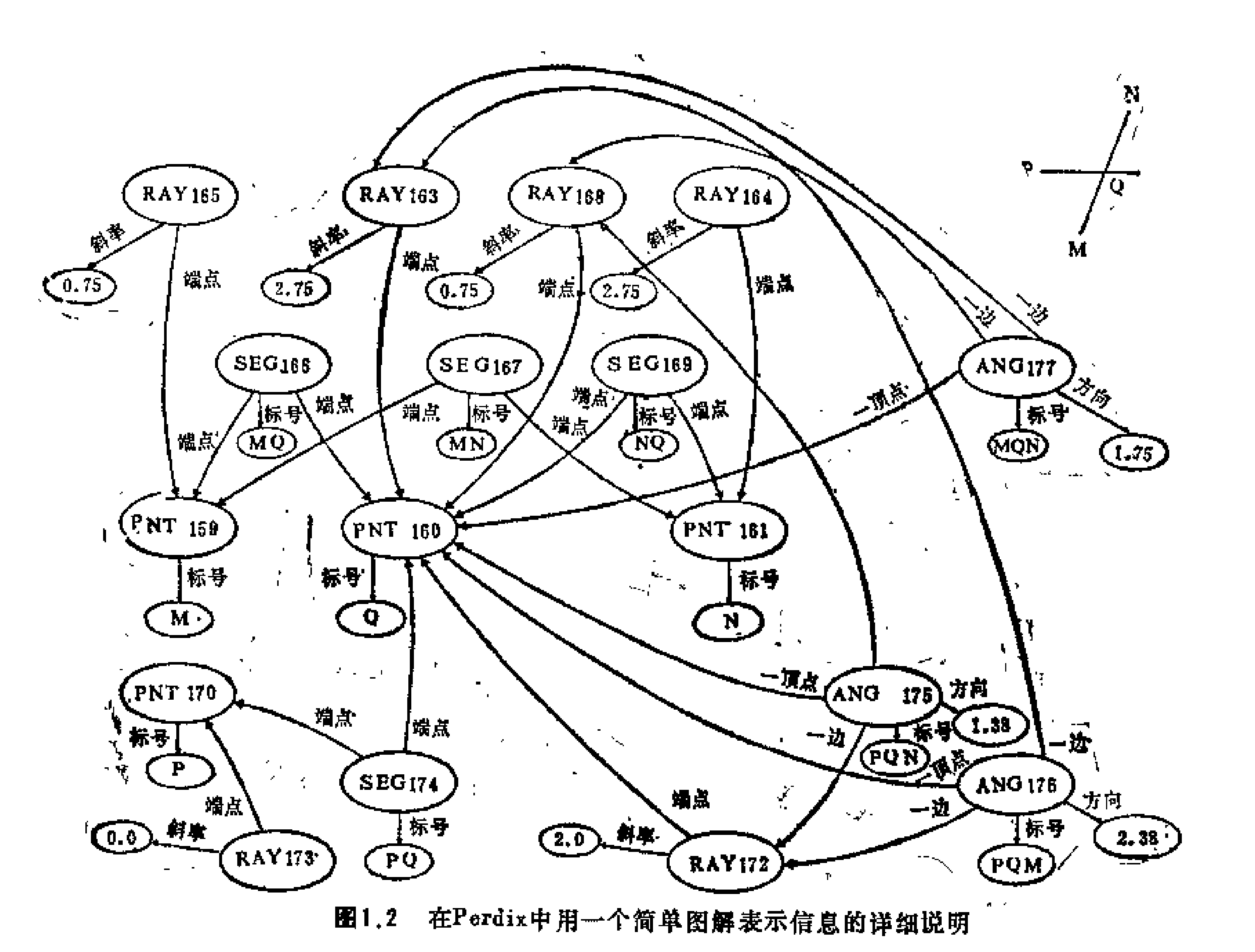


图1.2　在Perdix中用一个简单图解表示信息的详细说明

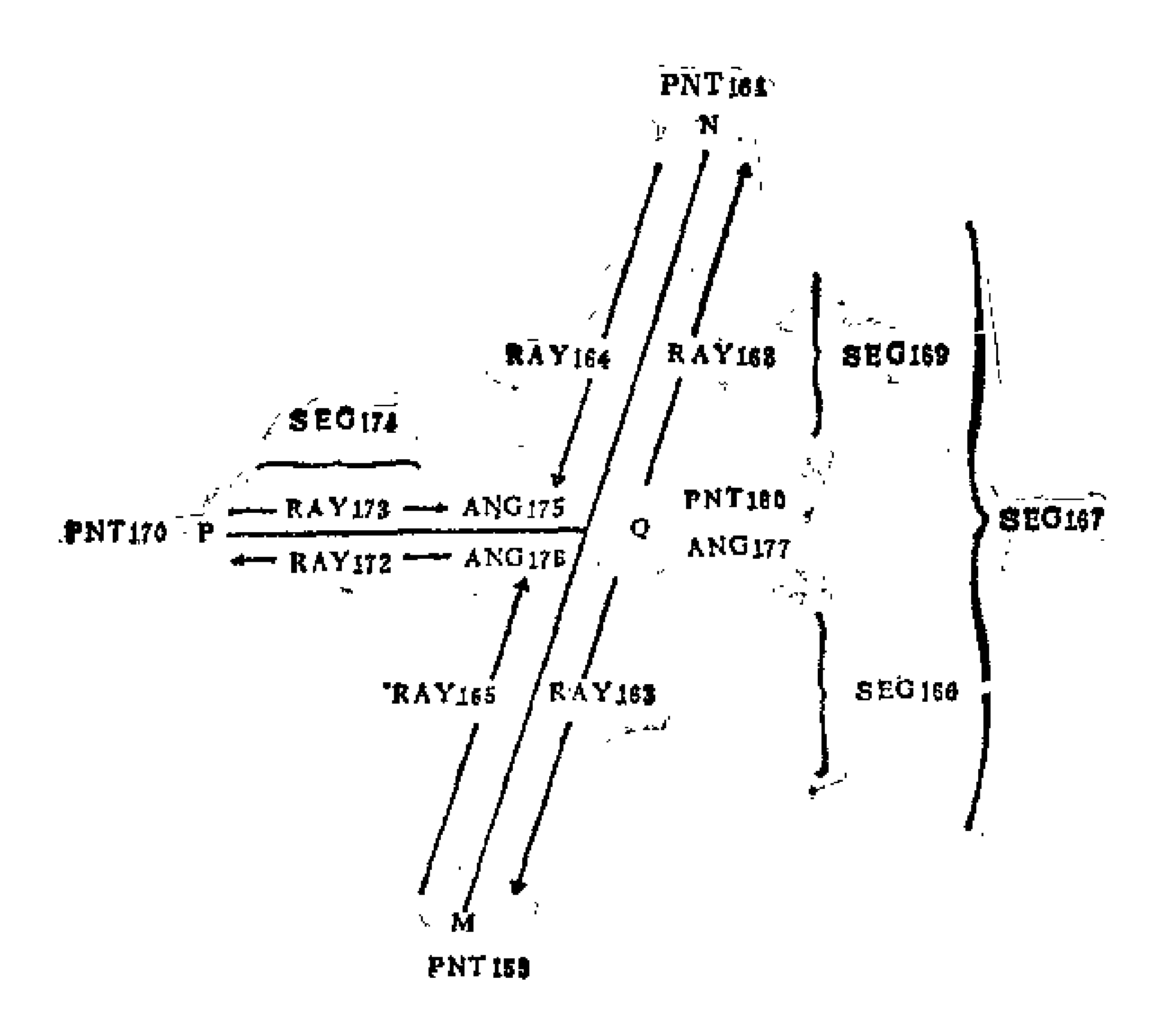


图1.3　图1.2所示图解的标号节点

每条射线的斜率可用一个类似极坐标的系统计算出来，而达到线性化时就可简化计算。一条向右的水平射线其斜率为0.0,一条向上的垂直射线其斜率为1.0,一条向左的水平射线其斜率为2.0,一条向下的垂直射线其斜率为3.0,居于这四条射线之间的一些射线，其斜率为0.0—3.0之间的值。表示每条射线的节点都与它的斜率和端点连接起来。因此，在图中RAY163是具有端点Q朝下的射线，而RAY168是具有端点Q朝上的射线。ANG175、ANG176和ANG177表示由相交的射线所形成的角。每个角都与它的顶点和作为它的两条边的射线相连接。作为每个角的方向的量就是角平分线的斜率。

为了使图形易于看懂，省略了两个重要的表示成分。在网络中有表示类别名称的节点：如点、线段、射线、直线和角。在网络中还包含了把这些类别名称与它们的例子连接起来的链。但在图1.2中无论是类别名称节点还是关系链都省略了。另外，还有许多表示部分—整体关系的链也没有表明。例如，RAY163、RAY164、RAY165和RAY168是直线M-N的各个组成部分，并且线段是相应射线的一部分，SEG166和SEG169都是SEG167的一部分，等等。

B.知觉模式识别的模拟

模式匹配是Perdix 全部作业的一个基本成分。模拟知觉模式识别的各种加工成分就是用Perdix的一般模式匹配过程来分析图形信息的那些成分。模式识别可以以两种方式进行。识别一个简单的模式只需要一个产生式条件就可以了，而识别较复杂的模式就需要一套相关的产生式条件。

用一个产生式规定的模式是一个含有某些规定的链的结构。例如，图1.4的(a)就表明了一个所谓ANGLEPARTS的模式。当产生式的条件包含这个模式时，这个系统就接受一个特定角的信息，例如，用这个模式来识别出哪个点是一个特定角的顶点，哪两条射线是这个角的二条边。在这个图形中，中间那个节点代表一个角，通过averx与顶点相连，通过aside与二边相连。

当对这个模式的一个条件进行检验时就有一个表示当时情境的网络搜索。如果情境中有一组能被指定到与那些指定节点关系一致模式的开型位置上，那么搜索就成功了。

当一个产生式条件包含一个模式时，这个模式的各个位置，即可以被规定为是一个常数也可以被规定为是一个变量，当检验这个模式时，有一些可以先赋值。现在假设Perdix要找出一个以一条特定线作为一条边的角。这个产生式的条件就如同图1.4(b)

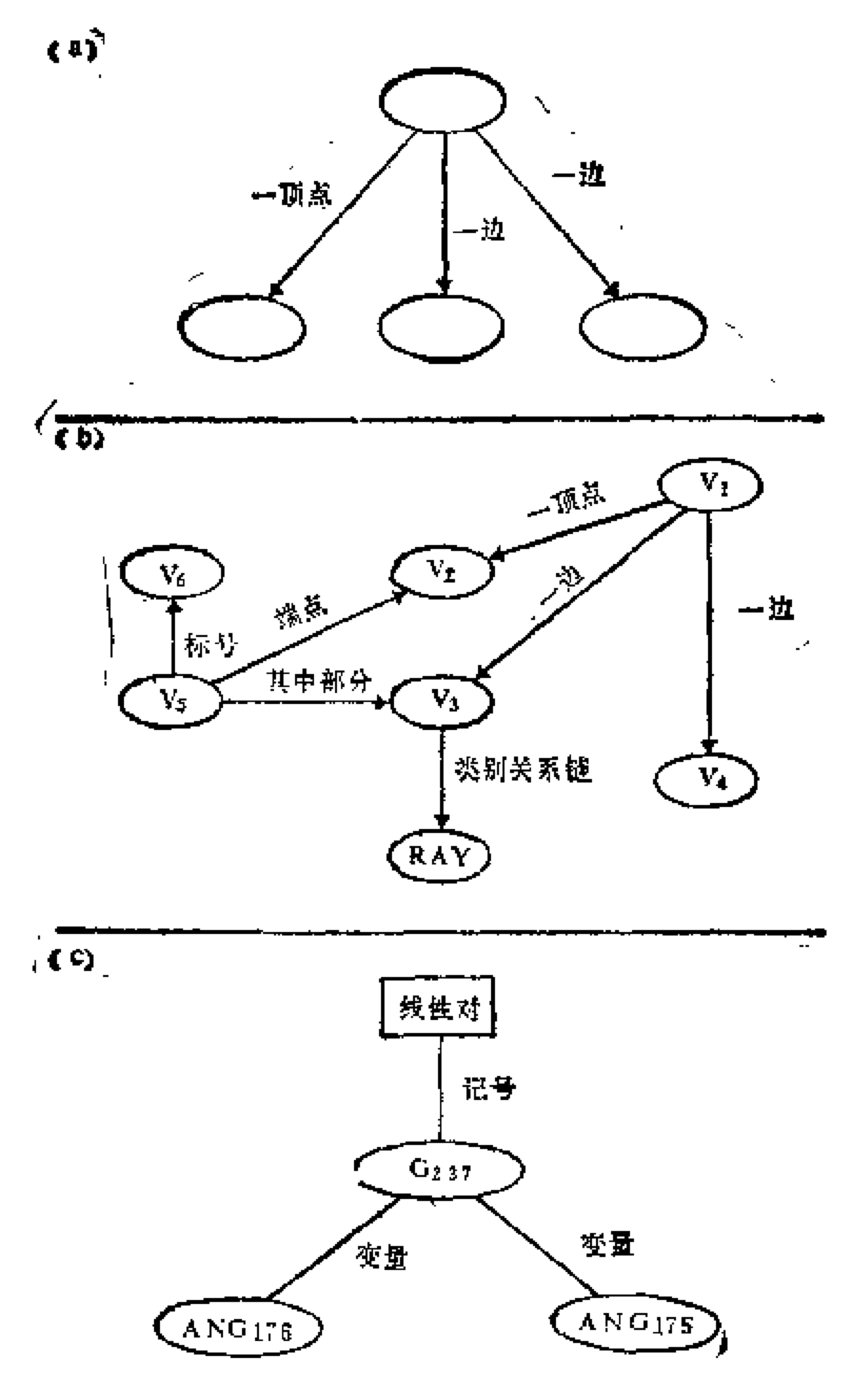


图1.4　在一个产生式中模式识别的举例

的模式，除了标号为RAY的节点是一个常数之外，它的所有位置都是变量。若要找出作为角的一条边的那条线段，可给变量V_6标上一个特定的标号。假设情境如图1.2所示，V_6的值是MQ。这就是说，为了满足这个条件，必须要有许多能被指定为V_1，V_2，

V_3，V_4和V_5，并与MQ和RAY以及相互之间都有直接关系的节点。正如在图1.2中所看到的，这个模式将被匹配，V_5将与标号为MQ的节点SEG166匹配。SEG166是RAY163的一部分。RAY163是属于射线类别，正如isa链所要求的，它表示类别关系。RAY163也和一个角有一边的关系。因此，V_3将与RAY163匹配。PNT160是SEG166的一个端点，它与一个角有顶点关系。PNT159也是SEG166的一个端点，但是PNT159不能满足顶点关系这个要求。因此，V_2将与PNT160匹配。可与V_1匹配的有两个节点，ANG176或ANG177，二者都以RAY163作为一条公共边，以PNT160作为顶点。这个系统将与其中之一匹配，究竟与哪一个匹配需要看初始表示时组成角的顺序。如果V_1与ANG176匹配，那么V_4就与其另外一边RAY172匹配。但是，如果V_1与ANG177匹配，那么V4就与RAY168匹配。

对于复杂模式识别，Perdix使用了相当于模式识别系统的若干套有关的产生式。对于平行线问题，就用一套产生式来找出每对角之间的各种关系：对顶角、补角、同位角、内错角以及同旁内角的关系。用判定网的方式把这些产生组织起来（费根鲍姆，1963；亨特、马林、斯通，1966）。当这个模式识别系统被使用时，就有一系列的检验。每次检验都要确定所识别的那个角需要的某一特点是否存在。下一步所要检验的特点依赖于前边检验的结果。如果找出了各种特点的一个适当组合，这个系统就识别出了它要知道的一种关系。

举一个例子，看图1.2中的∠PQM和∠PQN。这两个角有所谓的互补关系，而能决定这两角之间这种关系的那些特点是这二个角有同一顶点，有一公共边，而其余两条边是共线。如果Perdix具有分析∠PQM和∠PQN之间关系的目标，那么这个过程首先就要看出这两个角有同一顶点，再看出∠PQM的一条边也是∠PQN

的一条边，其余的两条边是共线，而这条共线和公共边并不互相垂直。最后，Perdix就把信息增加到表示这两个角之间互补关系的网络上。

组织一个产生式系统来识别模式的一种方式是基于特点判定树的形式(费根鲍姆，1963；亨特、马林、斯通，1966)。这种判定树的形式已用在Perdix中。图1.5就是一个例子。在这个图中，菱形代表所要进行的检验。如果检验是肯定的，就用标明"是"(Yes)的直线指向下一步。矩形代表Perdix所要执行的动作而不是检验特点。

这些产生式所执行的两种动作一是给变量赋值，二是把各种关系增加到表示情境的网络上。给一个变量赋值仅指给变量一个值并把这所赋的值存贮在记忆中。图1.5所示的判定树能识别出有同一顶点的两个角之间可能有的六种不同的几何关系。其中最常有的是对顶角关系，用VERT表示，以及线性对关系，用LINPR表示。这个系统还能识别出由两个角联合起来形成的一个完整的圆(用CIRCPR表示)；两个相邻的直角（这是LINPR的一个特例，用BOTHRIGHT表示)；平分一个角所形成的两个邻角(用BISECTPR表示)；或共同组成一个直角的两个邻角(用RIGHTPR表示)。Perdix还包括另一种特点判定树检验，可用来分析两条平行线被第三条直线所截构成的若干对角，如同位角、内错角、同旁内角。

为了说明这个模式识别系统，我要说一下对图1.2所示的∠PQM和∠PQN的分析。这个过程首先从表示其值为变量V_1和V_{11}的两个角的节点开始。ANG175赋值为V_1，这个角的标号是PQM，ANG175赋值为V_{11}，这个角的标号是PQN。使用这套产生式的判定已由Perdix所具有的计划过程所提供，并在这个阶段指定V_2为公共顶点PNT160的值。

在图1.5中用T_1表示所要试的第一个产生式。这个产生式有

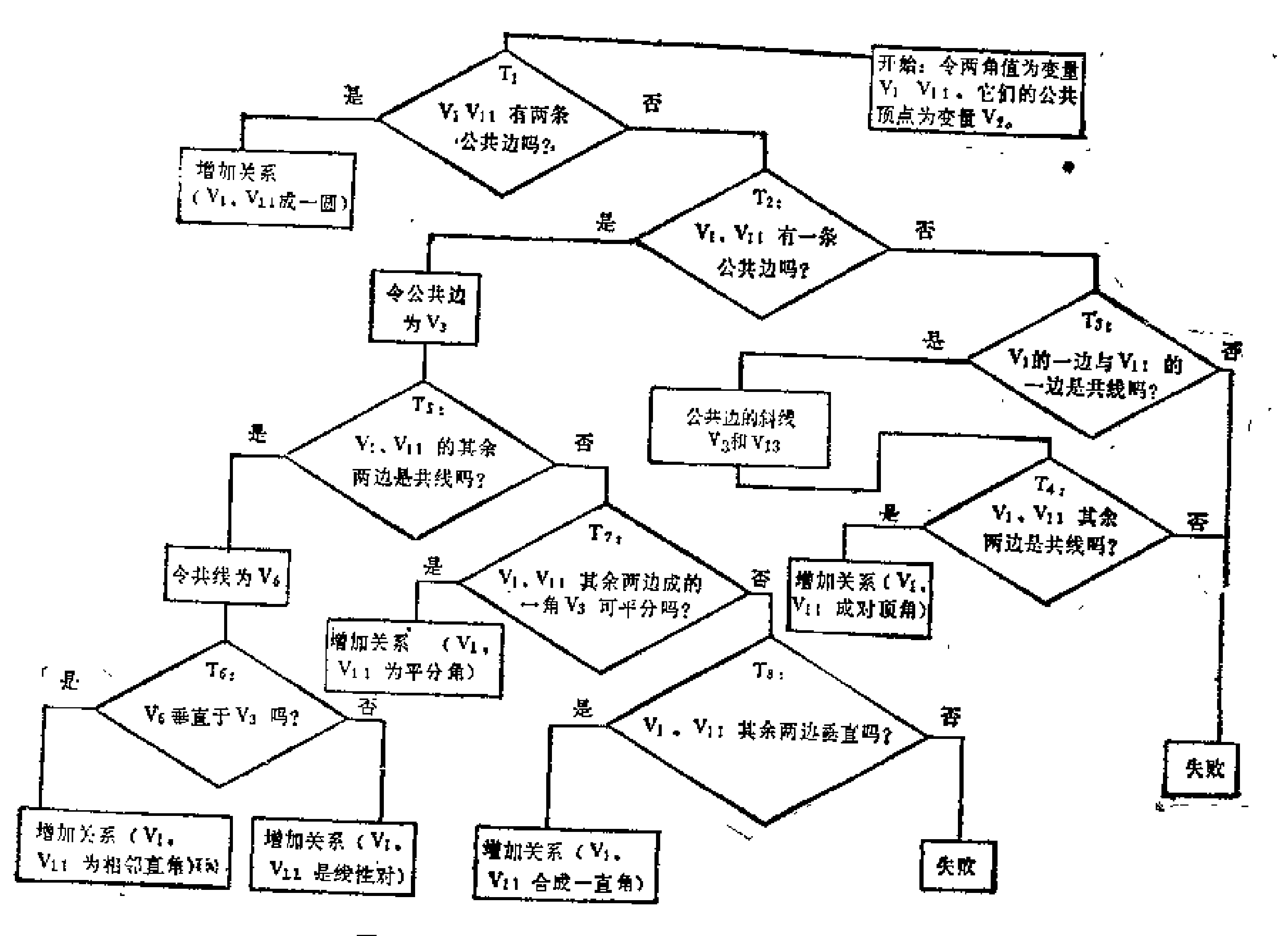


图1.5 分析同顶角所用的模式识别系统

一个条件，即要求V_1的两条边也同样是V_{11}的两条边。这个条件失败了，因为ANG176的两条边是RAY163和RAY172，而ANG175的二条边是RAY168和RAY172。下一个所要试的产生式是T_2。如果V_1与V_{11}有一条公共边，那么T_2的条件就是真的了。这个产生式的检验成功了，因为RAY172是两个角的公共边，并且RAY172赋值为V_3。

接着试的产生式的是T_5。T_5的条件是检验RAY163和RAY168，也就是这两个角的其余的两条边。检验这两条边是否同在一条直线上。刚好这两条边同在LIN162上。因此，T_5检验成功，LIN162赋值为V_6。接下去试T_6，这包括检验LIN162是否与包含RAY172的LIN171垂直。这个条件没有得到满足，因此最后的动作是把L INPR的关系增加到表示这种情境的网络上。

把表示关系的信息增加到网络上的动作表示了问题解决过程中进行推理的基本机制。图1.4的(c)表明了被增加作为∠PQM和∠PQN模式识别分析结果的表示关系结构。这个系统生成一个特殊记号的节点并与其关系的名称LINPR相连接。在图1.4中把这个特殊记号节点称为G237，通过标号arg(表示关系的自变量)的链与表示有这种关系的两个角的节点相连接。以后，通过其条件是要求详述两角之间各种关系的产生式就能识别出这种关系的存在。

V. 用于推理的定理

解几何题所需要的第二种一般知识是推理知识。几何题推理是根据课堂明确教过的一般定理进行的。例如："同位角相等，""同一条直线上的两个角互补"以及"一个三角形的内角和等于180°"。象前一部分一样，我还是先谈谈在Perdix中定理信息的表示。

A. 定理的表示

除了几何图形提供的信息之外，还有以文字定理的形式提供的信息，如"角P是40°，"或"直线a与直线b平行"。

在Perdix中，以文字定理形式提供的信息比起前面图1.1给的信息表示得要稍加复杂些。差别在于给每个关系实例加上记号节点。所以需要记号节点，这是因为这个系统必须在同一对角的各种不同的关系之间进行区分。例如，"对顶角相等"这个定理是说如果两个角能用关系VERT联系起来，那么这两个角也能用关系CONG联系起来。在Perdix中推理定理的表示是很简单的：每个定理都被表示为一个产生式。这个产生式直接采用了推理产生式的"如果……那么……"这种特点。在每一种情况下，产生式的条件就是去检验定理的前提是否存在，如果前提存在，就可执行产生式，动作是由增加的关系组成的，而这种关系是以定理的结论表述的。在图1.6中举了一个例子。图中V_1、V_2、V_3和V_4是变量。如果左边这个模式被得到了，那么在新结构的自变量位置上使用与已有结构的自变量位置上相同的节点就可构成右边的模式。

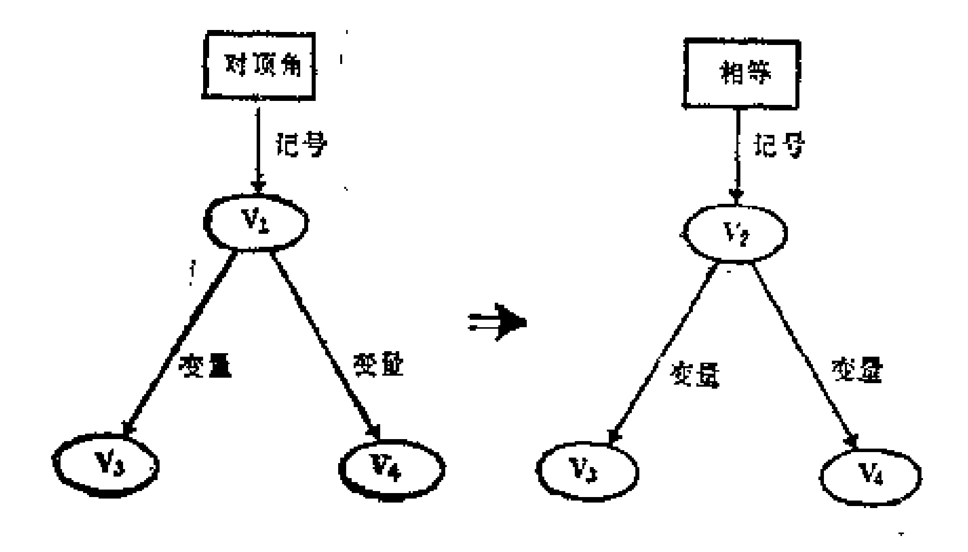


图1.6　定理的表示

B. 定理推理的模拟

进行定理推理所用的知识正是通常我们考虑作为几何课内容的那些信息，现在Perdix包括了有关平行线问题和全等三角形问题所有的主要定理的表示。

几何定理一般有两种。一种定理是推论出象全等或互补这种定量关系。另一种定理是推论出一个数字量度。

推出定量关系的定理，其前提要么含有几何关系要么含有其它定量关系。例如，在图1.6中，这个推理的前提是对顶角关系。含有定量关系的一个例子是一个产生式，这个产生式的前提是有两个相等的角，其中一个角与第三个角有某种定量关系。那么就能推论出另一个角也与第三个角有这种定量关系。例如，∠A和∠B相等，∠B和∠C互补，那么就可以推论出∠A和∠C互补。第三个例子是全等三角形相应的角或边也可被推论出是相等的。

许多推论出角的度数的定理都使用了象相等这样一些一般的定量关系。例如，如果∠A与∠B相等，并且∠A等于60°，那么∠B也等于60°。还有一些产生式是关于特殊角的，例如，一个直角是90°，一个三角形的内角和是180°。

从图1.1中可看到推理产生式的运用。用双线箭头表示推理。在(e)中，推论出A_1和A_8相等，前提是这两个角是对顶角。在(f)和(g)中也看到了类似的推理。在(h)中看到可从一种定量关系推出另一种定量关系：例如，由于A_{12}和A_8互补，并且A_8和A_6相等，所以A_{12}和A_8也互补。在(h)中还表明了具体度数的推理。通过一个推理产生式而增加的最后结果是A_{12}等于140°。为执行这个产生式所要找出的条件是A_1和A_{12}互补，并且A_1是40°。

表1.4所给的信息也说明了推理产生式的运用。这个表表明了每个被试在解这道几何题的过程中所推出的各种关系的顺序。

Ⅵ. 策略知识

在解一道几何题时，每做一步都会推出某种新的关系，这种推理或者根据模式识别，或者根据问题中提供的其它关系。在Ⅳ、Ⅴ两部分谈到了Perdix 进行每步推理的过程。但是这些知识成分只能使这个系统进行个别步骤的推理。此外，还必须具有能把每个个别的解题活动组织起来形成一个完整的解答的知识。

在Perdix中有二种密切相关的功能能把问题解决组织起来。这两种功能就是提出目标和制定计划。关于提出目标和制定计划的知识正是一个学生通过几何课教学所要获得的基本知识。这是一种技能方面的知识，可以使学生知道如何运用几何信息。

在这部分我从叙述Perdix提出目标和制定计划的基本过程开始。然后我再谈三个重要的问题，这三个问题都涉及到与问题解决的一般理论有关的几何策略知识的性质问题。第一个问题是一个比在第Ⅱ部分所说的系统更加复杂和灵活的目标结构。第二个问题是问题解决定势和问题解决期间选择计划的过程。第三个问题是解几何题的结构和在问题中最初给的问题空间的完备性。

A.提出目标和制定计划的模拟

Perdix总的有两类不同的目标，它们涉及到不同情境所要求的不同动作，在这两类目标中，一类目标是为了证明什么东西，另一类目标是为了发现什么东西(参看波利亚1962、1965)。一个目标结构还包括达到目标时那种情境所产生的一个模式。当要达到的目标是证明什么东西的时候，这个模式就是表示被证明的定理的关系结构。如果动作是找出什么东西，那么这个模式就是不完全的，既要指明已有的成分又要指明完备这个模式所需要的那种

成分。举一个例子说明，假定我们当前的目标是证明两个角相等，这个目标的动作就是要证明什么东西而不是发现什么东西，那么这个目标的模式就是两个指定角之间的相等关系。再举一个例子，假定目标是找出两个角之间的一种定量关系，那么这个目标的动作是要发现什么东西，这个模式是一个不完全的结构，包括两个角和一个需要某种定量关系的指示。

Perdix制定计划的知识在某些方面与萨塞多蒂(1975)开发的NOAH系统类似。萨塞多蒂提出了一个把各种动作知识存贮在所谓过程网络结构中的概念。所存贮的每个动作信息包括执行那个动作所需要的必要情境特点和执行那个动作所产生的结果。就是用这种信息来构成一个解题的计划，在这个解题计划中所包括的各种动作，其结果与后来必须执行的动作的先决条件是一致的。

在Perdix中，制定计划的信息与目标结构的各种不同的模式相联系。当目标模式涉及一个角的度数时，可行的计划是使用一个三角形的内角和计算出所要求的度数或者使用未知角与某个度数已知的角所具有的某种定量关系计算出未知角的度数。当目标模式涉及到两个角之间的定量关系时，可行的计划是根据关于旋转的几何关系推出所需要的关系，根据两条平行线与一条截断线的几何关系推出所需要的关系，证明两个含有目标角的三角形全等，或者使用其它角形成一连串定量关系。

Perdix所知道的每个计划都有若干先决条件，并且要加以选择，一个计划的各种先决条件都必须在情境中出现。先决条件的特点不足以保证计划生效，但对于审查这个计划来说有足够的价值，可使计划有希望实现。例如，要根据两个角之间的旋转关系推出一种定量关系，先决条件就是这两个角必须有一个公共顶点。对于要通过证明两个三角形的全等来证明两个角相等的计划来说，其先决条件是所要证明的这两个角必须要在这两个三角形里。

当解题期间提出了一个目标时，就要执行一套关于那个目标的计划产生式。在这些计划产生式中，条件就相当于为那个目标可用的各种计划的先决条件。每个产生式的动作就相当于采纳一个指定计划的决定。如此的结果是所采纳的计划其先决条件必定得到满足，首先看到哪个计划的先决条件就采纳哪个计划。

在Perdix中，仍用表示模式和定理信息的那种一般网络结构来表示目标和计划。图1.7就是一个例子，目标结构有一个标识记号节点G238，它与这个结构中的四个节点相连接。其中包括目标所涉及的动作种类，所要达到的目标模式类别，和这个目标模式中所包括的各种成分。在这个例子中，其动作是找出一种关系，用XFIND表示，目标模式是某种没有指明的定量关系，用?QREL表示，成分是两个角，ANG175和ANG176。

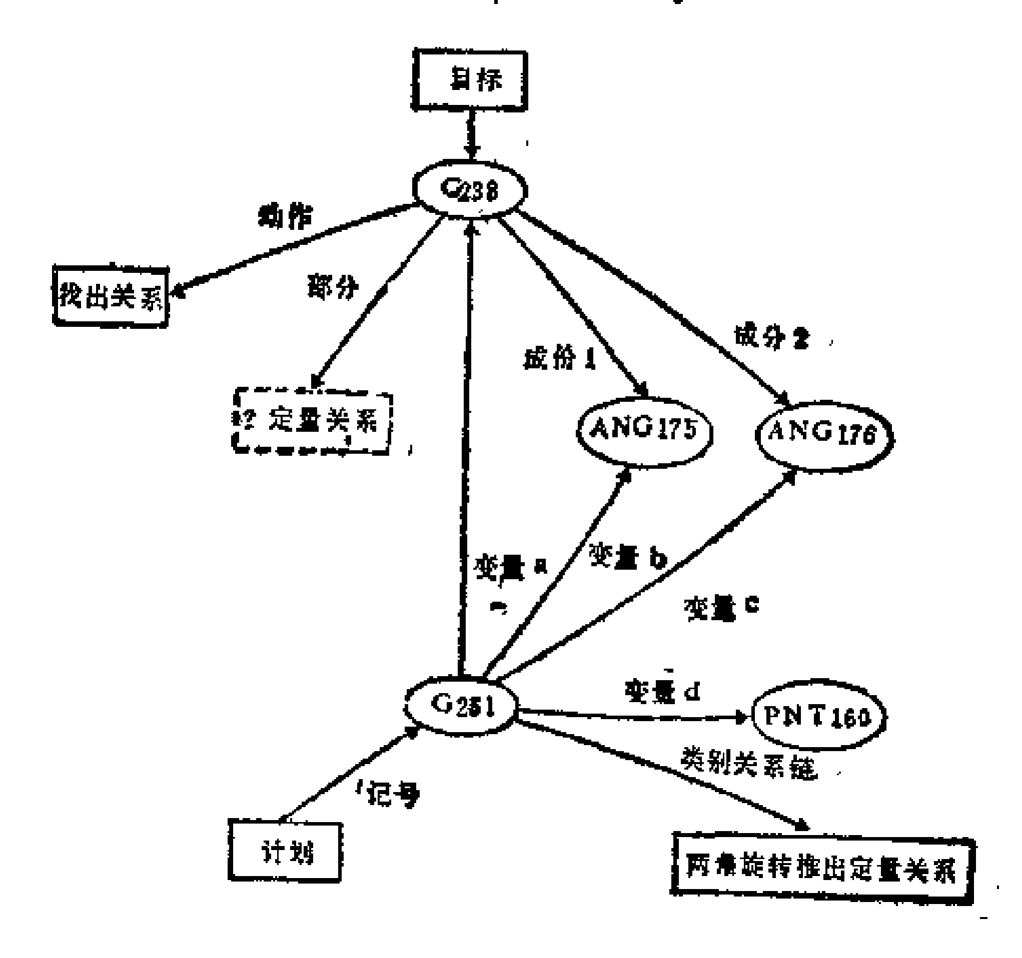


图1.7 在Perdix中目标和计划的表示

计划结构使用表示关系命题的一般形式来表示。一个特定的计划是用一个记号节点G251来代表的。这个计划企图根据两个角之间的旋转关系来推出一种定量关系；这个比较复杂的类别名称被缩写成RTNRELQREL。在表示关系结构中联结起来的节点包括识别计划所要达到的目标的节点，目标达到时所连接起来的两个角和另一个与计划有关的目标PNT160。

在问题解决过程中，当Perdix采纳了一个计划时，就允许立即执行这个计划。但是在大多数情况下，还要做进一步的工作，这就要求Perdix提出一个子目标。子目标的知识对于实现计划来说是很有用的，正是这些关于子目标的知识构成了Perdix计划知识的一部分。通常所要求的子目标是明确的。例如，如果ANG175和ANG176由对顶角或互补角这种旋转关系连接起来，那么Perdix就能执行图1.7所示的计划。如果情境中没有表示出所需要的这种关系，那么就要提出子目标，通过知觉模式识别去试着找出这种关系。随时注意用新的子目标代替旧的子目标，但是当新的子目标已经达到或不可能达到时，存贮在记忆中的旧的子目标又恢复起来。

为了说明提出目标和制定计划的过程，再回过头来看图1.1。在题目中所给的目标作为一个主要目标。方框(c)所示的目标结构象图1.7所示的结构一样是一个简略的图示，其中的动作是要找出什么东西，而模式是用一个未知数表示的角A_{12}的度数。为了求出度数，Perdix检验了两个计划的先决条件，一个计划是使用三角形的内角和，但在这个题目中是不可行的，因为所给的几何图形中没有三角形。由于A_1是已知角，于是Perdix采取了关于与A_1有定量关系的计划，求A_{12}的度数，并且提出了找出这个度数的目标。把求A_{12}的度数这个目标存贮在记忆中，而找出A_{12}与A_1的定量关系这个新目标就成为当前的目标。

要找出一种定量关系就要制定计划，而Perdix就要检验其可行计划的先决条件。第一个计划涉及到三角形的内角，这在本题行不通。第二个计划要求两个角有同一顶点；这也不现实。第三个计划要求这两个角有两条边是平行的，另两条边在一条截断线上。A_1和A_{12}有两条边是平行的，但是另两条边不在同一条截断线上，因此这个计划也不可行。此时，Perdix就采取了这样一个计划，用几个与A_1相等的链接角去把A_1和A_{12}联系起来。于是Perdix就确定了新目标，找出与A_1相等的角，而把找出A_1和A_{12}之间的定量关系这个目标存贮起来。这就是图1.1方框(d)所示的情境。

要找出几个相等的链接角受着知觉扫描过程的支配。把A_1和A_{12}的顶点连起来就形成一条扫描线，按其离扫描线距离的远近，按次序考虑每个角，离扫描线近的角先考虑，离扫描线远的角后考虑。A_{11}和A_6都正好在扫描线上，但考虑这两角的顺序也有先有后，这依赖于所给图形的偶然因素。如果先考虑A_{11}就构不成一个可行的解题计划，因为A_{11}和A_1之间没有所要求的几何关系。而先考虑A_6，这个计划就可行了，因为A_1和A_6有一个公共顶点。然后进行模式识别，发现A_1和A_6是对顶角，就可推出A_1和A_6相等，这正是我们所需要的。这样就找到了第一个链接角。

由于找出链接角这个子目标已达到，于是Perdix就确定下一个子目标，找出这个链接角与未知角A_{12}之间的定量关系。但是正如不能把A_1和A_{12}直接连起来一样也不能直接把A_6与A_{12}连起来。仍需再确定一个子目标，再找出一个链接角，其情境如图1.1的方框(e)所示。

要找出A_6和A_{12}的链接角，象在前面一样也使用了上述的扫描线。A_{11}是扫描线上的一个角，但是不能以任何所要求的方式把它与A_6联系起来。(注意，正是在这一步，表1.4中的被试4把A_6

和A_{11}之间的关系看错了。这个学生的计划过程显然包括一套比他们应当具有的更加一般的必要特点。而且，这个学生对于省略了下一步应证明所包含的特定关系这个计划结果充满了信心。)A_7和A_{10}都同样靠近A_6和A_{11}* 之间的这条扫描线，并且满足了涉及平行线关系的先决条件，但是，当检验这两个角时，发现它们与A_6并不相等，而Perdix要求找出一个与A_6相等的链接角。其次，A_3、A_6、A_8和A_{14}也都离扫描线距离相等，这四个角中先考虑哪一个，哪一个就是下一个链接角，因为它们与A_6或者是同位角或者是内错角，所以都与A_6相等。而在本题中恰好首先考虑的是A_8，因此接着要提出的子目标是找出A_8和A_{12}之间的关系。

当根据定量关系的计划产生式来考虑A_8和A_{12}时，发现A_8和A_{12}满足了有两条边平行，另外两条边在同一截断线上这个先决条件。这就使得Perdix采纳了PRLRELQREL这个计划。根据平行边的关系推出定量关系。把要找出所需要的几何关系做为一个子目标，这就构成了方框(f)所示的情境。

如果存在着平行线的关系，那么通过模式识别系统就可以找出这种关系。执行了这种过程就找出了同旁内角的关系。由于这个子目标已达到，Perdix就转到下一个目标，找出A_8和A_{12}之间的定量关系。现在要执行的计划是根据平行边的关系进行推理，得出了A_8和A_{12}是互补角的关系，其情境如方框(g)所示。

方框(g)还表明，保留在记忆中的目标是找出A_6和A_{12}之间的定量关系，进一步说是A_1和A_{12}之间的定量关系，也就是A_{12}的度数。那么第一个子目标可以达到是因为A_6和A_8相等，同时A_8和A_{12}互补，所以可推出A_6和A_{12}互补。当做出这个推论时，就能达到下一个子目标，推出A_1和A_{12}也互补。最后，Perdix回到问题的主要目标和达到这个主要目标的计划上。Perdix执行了上述计划得

* 原文为A_{10}，疑误。

出了答案。

不确定的目标结构。上述目标结构是最流行的问题解决理论所涉及到的典型的目标结构。在问题解决过程中，问题解决者随时都有一个目标，这个目标是由情境特点特别组合而成的，而问题解决活动就是去构成这些特点的组合。

在下边几节我要对这种理论表示作一发展，提出不确定目标的表示，这种不确定的目标可由几个任选的特点组合来满足。真实的人类几何问题解决理论要求做这种发展。因为解决一些重要的问题正是使用了这种更加灵活的目标结构。中学生解几何题的记录材料表明确实需要这种被发展的目标表示理论。下面我们就说明这种情况。

证明三角形全等这类问题就使用了不确定的目标结构。看图1.8，三角形中所画的细线和右上角所列的表是学生解题所做的工作。要证这道题，就要用$\overline{PM}\perp\overline{QR}$来表明∠PMQ=∠PMR，用$\overline{PM}$平分∠QPR来表明∠QPM=∠RPM，并且注意到$\overline{PM}=\overline{PM}$，这样通过上述角-边-角的关系，证明了△PQM≌△PRM。

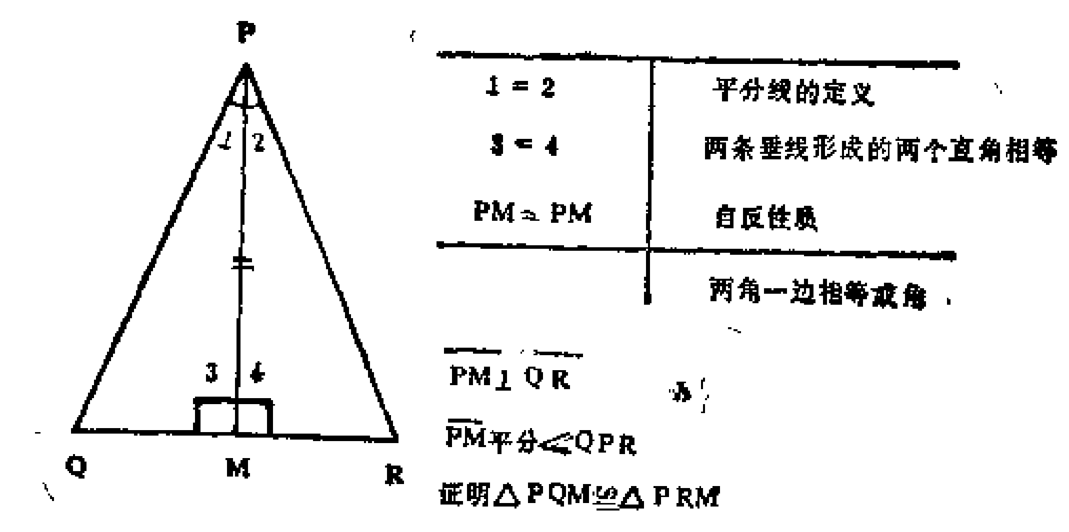


图1.8　表1.5所述的那个学生在求解过程中所添加的辅助线和所写出的证明步骤

有几种不同的途径可证明三角形全等。除了上边用过的角-边-角的模式外还可以用边-边-边关系，边-角-边关系，角-角-边关系，以及直角三角形的斜边是否相等这些模式来证明三角形全等。

对于象Perdix这样一个系统来说很可能使用图1.7所示的目标结构来证明三角形全等这类问题。为此，当主要目标是证明三角形全等时，就要把这个系统的程序编写成其子目标是生成一个全等模式。首先可以试边-边-边这种模式，当不能证明三个边相等时，再试边-角-边模式，这样继续下去，直到有一种模式被证明成功为止。经由特定子目标进行反向工作的策略对Perdix程序来说是比较容易做的，而且还把这种策略用于解几何题的人工智能系统上(盖伦特，1963；戈德斯坦，1973；乌尔曼，1975)。没有特别编写解几何题的一般问题解决者(恩斯特和纽厄尔，1969；纽厄尔和塞蒙，1972)，但一般问题解决者达到目标的步骤最适用于确定的子目标结构形式，因此类似于已建立的几何问题解决者。

表1.5是一个学生解图1.8这道几何题的记录。解题过程表明，在找出一种足以证明全等的模式之前没有使用特定的子目标。这是一个学生解几何题时我与他交谈的谈话记录。

表1.5　　学生解图1.8问题的谈话记录

*1	S：	对，PM垂直于QR，而且PM平分∠QPR，所以这两个角相等，证明了这两个三角形全等。
*2		这就是说这两个角相等，因为这条垂线平分了这个角。
	E：	嗯。
	S：	那么……你要让我写出正式的证明吗？
	E：	当然，至少是一部分证明。
	S：	好，那么……
*3		这两个直角所以是直角因为……如果我能很快地画出这条垂线……

*4 对，∠1和∠2

E：对。

S：∠1等于∠2，因为……PM平分了它。

E：对。

*5 S：如果我要写出理由……噢，这是已知的条件。我要想想我是否知道理由。对于平分。但是，当……角平分线的定义，我猜到了。

E：这很好。

*6 S：对，而且……还得出了∠3和∠4。∠3等于∠4，因为二条垂线构成了两个直角；而所有的直角是相等的。

*7 其次，PM等于PM，这是因为自反性质。……此外我还能用角-边-角。或者我能用另一种方法做，我可以说……我有了选择权。

E：确实如此。

S：我可以说……

E：你可对我说说选择权。

*8 S：好，我已说过角-边-角，或者，我还可以说……这是两个直角三角形，因为它们都包含有一个直角，那么我可得出斜边-角。

E：好，那么你能告诉我，你认为什么时候你要使用角-边-角呢？

*9 S：现在就能……这我知道……我正寻找我要证明相等的东西。一条公共边是相等的，平分线把一角分成两个相等的角，而且我知道这两个角是相等的，所以我就试着把它画出来。我试着证明什么相等，我看到什么，就写出什么。

E：我明白了，很好。

注：引自J.G.格里诺的"良结构问题中的不确定目标"("Indefinife Coals in Wele Structured Problems")载Psychological Review，1976，83、479、481。

在*1处，学生大声读已知条件。在*2和*3处，学生找出相等的部分。在*4处，学生标出图1.8的三角形中∠1和∠2的号码。在*5处，学生写出图1.8右上角所示的第一行证明。在*6和*7处，学生标出∠3和∠4的号码，并且完成了证明。在*8处，学生提出另一种证明，在*9处，学生说明所使用的解题步骤。

从表1.5的谈话记录中看不出学生在完成角-边-角这种模式之前使用了确定的子目标。如果有这种子目标早应在* 2处之前出现，那时学生可以说："我从找出角-边-角开始。"如果学生在* 7处的说明之前说："我需要找出相等的边来完成角-边-角的形式，"这也不大能直接证明他使用了这种子目标。这句话表明了学生先想出了这种模式，然后才看出相等的成分。

这个谈话记录实际上有两处相当明确地表现出这个学生没有把各种模式特别看作为子目标。从* 9处看出，这个学生相当强烈地感受到他是泛泛地找出相等的成分而不是寻找相应于某一假说的特定模式来解题的。另一处表明学生看出了一个已用模式的替代模式。在* 8处找到了这个模式的另一种解释，在* 9处做了详细说明。如果这个学生直接寻找某一模式所涉及的成分，这个学生就不大可能看出已找出的成分也适于其它模式。

关于表1.5的结论是有代表性的。我收集了一套中学生解几何题的谈话记录，有24份是关于证明三角形全等的题目（详看格里诺的文章，1976）。在这24份谈话记录中，只有一份记录明显地表明那个学生在开始解题时考虑到了一个确定的子目标。有三份记录明显地表明他们在找出成分对的完整模式之前考虑过指明的子目标。这就表明解决这类题目的典型方法是泛泛地寻找相等的成分，同时监督着这一过程的进行，当一个成分出现时，识别出成分对的完整模式。正如一个学生所说："我只是让这种材料，已知的信息有几分浸透我的头脑，……我喜欢考虑这种信息，并且只是把这种信息记在心上，你知道吗？处处无需做什么猜测，你懂吗？"

这些材料有力地支持了这个结论，在证明三角形全等这类问题中，学生主要是泛泛地寻找一对相等的成分，而不是根据某一具体模式的假设进行寻找。为了使Perdix能模拟学生作业的这个方面，极需要在理论上做一重大的发展。通过把目标结构表示为

能找出问题情境中的相等成分任一组合的模式识别系统来模拟泛泛地导找。

在第Ⅳ部分叙述了Perdix模式识别的一般形式。在象EPAM(费根鲍姆,1963)和CLS(亨特,马林和斯通,1966)这些最新的模式识别理论中所使用的规则都是标准的。表示目标的模式识别系统有许多要检验的特点，通过特点的几种不同的组合才能满足目标。图1.9表明在模式识别网络中大约有一半节点是用来证明三角形全等的。Perdix能识别出六种不同的模式,已说明了四种。在这个结构中有13个检验特点的节点。

为了弄清楚模式识别目标结构是怎样工作的，先假定问题解决者已找出一对相等的角和一条公共边$\overline{PM}$,如图1.8所示。现在看图1.9所示的第一个检验;这个检验是问在所要证明的两个三角形中是否有一对相等的角。确有这么一对相等的角，假定Perdix首先找出∠QPM和∠RPM,这两个角就成为变量V_2和V_{12}的值.现在Perdix又检验这两个角的边是否相等。$\overline{PM}$是∠QPM和∠RPM的公共边,由于$\overline{PM}$等于$\overline{PM}$自身,所以检验成功了,并且$\overline{PM}$成了变量V_5和V_{15}的值。依次进行的下一个检验是看∠QPM和∠RPM的另两条边$\overline{PQ}$和$\overline{PR}$是否相等;无法证明$\overline{PQ}$和$\overline{PR}$相等，因此这个检验失败了。再检验是否还有一对已知角也相等。结果发现∠PMQ和∠PMR相等，并且它们成了变量V_3和V_{13}的值。最后确定了$\overline{PM}$是V_5的现行值，它是∠PMQ的一条边，因此所找出的模式是角-边-角(ASA)。

要解图1.8中的这样一道题,Perdix使用了模式识别目标结构以及前边讨论过的模式识别、推理和确定目标的其它产生式。当Perdix开始解这道题时,知道总目标是证明△PQM和△PRM全等。这就导致了使用寻求相等的模式识别系统来进行检验。在这两个三角形中没有一对成分已知是相等的，因此Perdix就提出一个子

目标，在两个三角形中，找出一对相等的成分。

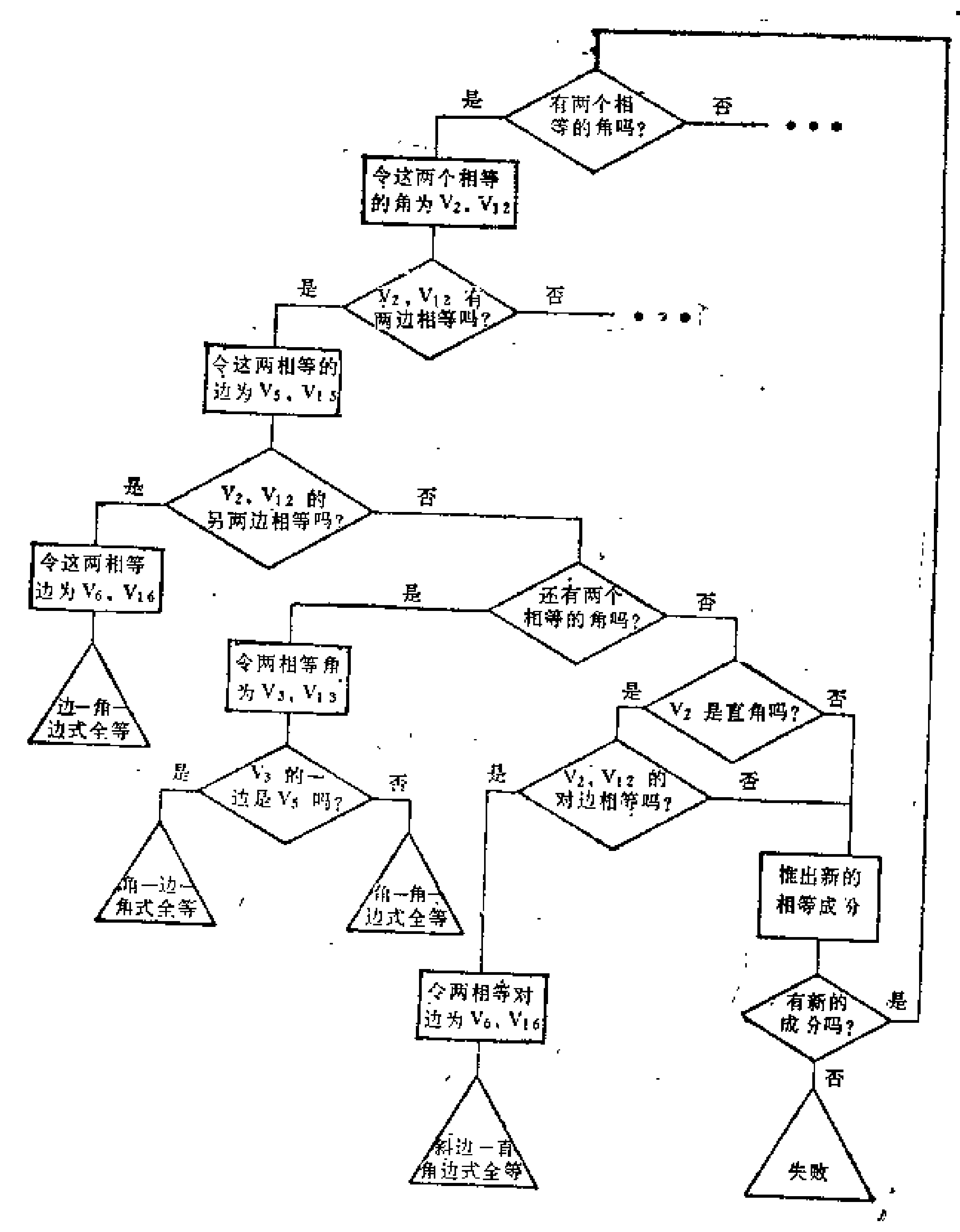


图1.9 Perdix证明三角形全等的目标结构，以模式识别系统来表示

推论出一对相等的成分是通过一套产生式完成的，在这套产

生式中其条件就是含有相等关系的特点模式，在解图1.8这道题时，成功执行的第一个产生式就是找出$\overline{PM}$是两个三角形的公共边，在这种情况下，Perdix把相等关系添到数据结构上，并自反地运用到标号$\overline{PM}$的线段上。然后Perdix回到主要目标上，再一次运用模式识别目标结构检验相等。只有两条边相等这一种关系是不够的，因此还要提出另一个子目标，推出其它相等的成分。这次所执行的产生式其条件是平分一个角，通过平分线所形成的两个角都在目标三角形中。而所执行的动作是推出∠QPM等于∠RPM。在回到主要目标之前，Perdix还要再一次检验一个蕴含三角形全等的模式，但是只有这些相等的成分还不够。还要再执行检验相等成分的产生式，这次找出了垂直线，由此推出∠PMQ等于∠PMR。现在再进行模式识别，就有足够的相等成分来确保识别出角-边-角这种模式，问题的目标达到了。

B.逐级计划和问题解决定势

我们常常看到人类问题解决者不能顺利地解题，识别不出能使顺利解题成为可能的某种情境特点。卢钦斯(Luchius,1942)报告了几个实验结果，在实验中，有一系列的问题都使用一种解决方法，那么被试就沿用这种方法去解另一个相比之下更为简单的问题。卢钦斯的实验所使用的问题大多数是关于水瓶问题，要把几个瓶中的水倒来倒去，最终使一个瓶中剩有规定数量的水。在这些大家熟知的实验中，由于反复使用一种方法而得出了一个公式：先把第二瓶装满水，再从第二瓶中往外倒水，倒出的水恰好足有装满一个第一瓶和两个第三瓶那么多，而第三瓶所剩下的水就是所需要的规定数量的水了。接着再给被试一个问题，这个问题既可以用上述公式解也可以用比较简单的方法解：先把第一瓶装满水，然后倒出够装满第三瓶的水，第一瓶中所剩下的水就是所需要的

规定数量的水了。大多数被试都使用了上述惯用的比较复杂的方法，而看不出来可以使用后来所说的那种比较简单的方法。对于这个可用简单容易的方法解决而不必用上述公式解决的问题，许多学生完全束手无策。

卢钦斯还扼要地报告了几何问题解决小型研究的结果。他连续出了四道题，在每道题中他都要求学生通过证明三角形的全等来证明所含内角是相等的。接着卢钦斯再出一道题，这道题既可用三角形全等的方法证明，也可以用对顶角相等的简单方法证明。10个学生中，有8个学生用三角形全等的方法证明。此后，卢钦斯又出了一道题，不能用证明两个三角形全等但可根据对顶角相等的定理来解题。对于这道题，10个学生中有六个在规定的2.5分钟之内没有做出。相反，没有做过上述具有不适当计划的问题的12个学生全都很容易地做出了这道题。

Perdix所使用的这种制定计划的系统可为问题解决定势提供一个证明。正如在这个部分开头所说的那样，Perdix所模拟的制定计划的知识包括一套设法解题的替代途径。当提出了某一目标的时候，Perdix就检验具有总体特点的情境，并选择一个具备适当先决条件的适当计划。只要采纳了一个计划，Perdix就执行与计划相应的活动，直到计划执行完了，或者断定计划是不可行的时候为止。

Perdix计划过程的设计有自顶向下和自底向上信息加工之分。如果自顶向下进行信息加工就有一个指导系统活动的总的概念和假设。如果自底向上进行信息加工系统的活动就没有特定的方向或预想，完全是根据情境中的信息做出决策的。实际的信息加工系统决不是只按一个方向进行信息加工的，或者仅是自顶向下，或者仅是自底向上，但从总的概念和设想而不是从问题情境的特定信息来看，各种系统在控制范围上确实是不同的。

当Perdix提出了一个目标时，就按着自底向上这一限定的方式制定计划。一个计划的选择依赖于对成功地执行计划来说是必需用的那些情境特点的检验。因此选择一个计划也许更多地依赖于情境信息。一个完全自顶向下的计划系统只按固定顺序来检验计划，而不管情境的特点。而另一方面，计划过程只考虑与Perdix所包含的替代计划有关的那些特点，因此计划系统不完全是自底向上的。进一步说，对特点的检验是连续进行的，每次只检验一项计划的先决条件。在第一个计划里，只要其中适当的先决条件被发现，那么这个计划就被采纳了。采纳了一项计划之后，Perdix就坚决地按照自顶向下的方式运行，一步步地实现与计划有关的子目标。Perdix虽有办法确定现行的计划不能成功，但是一旦计划被采纳了，就要花很大的气力才能废弃这个计划。如果能成功地执行一个计划，Perdix就总是按这个计划运行，直到最终地实现计划为止，而不会中途变卦另找简单或有效的计划。

Perdix解卢钦斯设计的几何题的作业依赖于对计划先决条件检验的顺序。在卢钦斯的几何题中，对Perdix可采用的两项计划给出了先决条件特点。如果需要证明相等的角是三角形的 内 角，那么这正是证明三角形全等这项计划所要检验的特点。如果两个角之间有一个公共的顶点，那么这正是找出旋转关系的那项计划所要检验的特点。选择那一项计划完全依赖于首先检验哪一项。我们可以预料，在大多数情况下首先检验是否有公共顶点，因为这或许是一个相当醒目的特点。但是，卢钦斯的实验结果认为，采用某一计划的最新体验能使得这一计划的先决条件特点变得更加鲜明，因此，即使别的计划更简单，Perdix也倾向于选择这一计划而不选择别的计划。我们没有把改变检验产生式顺序的 机 制 编 进Perdix程序，因此Perdix不能提供关于形成问题解决定势过程 的理论。但是Perdix计划产生式的各种安排说明了产生问题解决定

势的学生与没有产生问题解决定势的学生是如何的不同，而这种差别似乎造成了定势体验结果的不同。

C.需要作图的问题

要解几何题，有时必须增加一条辅助线或进行其它作图；虽然在几何问题解决中大多数作图是很简单的，但是却给问题解决理论提出了一个有趣的问题。大多数问题解决理论，包括现在所说的这种Perdix形式在内都主张，要解决一个问题，需要对这个问题已可用的材料进行变换，当一道需要作图的几何题缺少一个必要的成分时，学生就必须自己加上这个成分才能做出这道题。

有些题需要作图，这与纽厄尔和塞蒙(1972)提出的问题空间这个概念有关。问题空间是由问题所提供的材料、问题解决者所具有的改变情境的方法，问题解决者在选择计划和问题解决运算时所用的知识以及问题解决者在解题过程中所创造的情境组成的。在一道几何题中，问题空间的一个重要组成部分是问题中所给的图形。这个图形可以提供一系列几何要素——点、线段、角、三角形等等，这是解题时需要分析的一些要素。如果一道题不需要作图，那么问题空间开始所给的要素就足够了。要解这道题，学生必须在各种要素之间建立起一系列的联系，包括许多结构推理，学生不需要增加新的几何要素。而对于一道需要作图的问题，学生必须增加所需的几何要素。

根据现有的问题解决理论，我们还不能完全理解需要作图的问题。一个主要的理由是大多数理论都把问题解决看成是解的搜索，如果再加上作图，那么搜索的困难就更大了。如果情境中的要素是固定的，那么可能的动作数可随时限定到按问题所给的规则重新安排要素组合方法的数目上。但是，如果有可能往情境中增加要素，那么就可因为有许多不同要素可增加而使得可能的动作

数变得很多。因此在大多数系统中不包括作图这个过程，这样一来，需要作图的问题就成为现有的系统所不能解决的问题(如乌尔曼，1975)。

而另一方面，学生都能很容易地解答各种需要作图的题。有一道例题是证明等腰三角形二底角相等的定理，既证明如果一个三角形的两腰相等，那么两腰所对应的两底角也相等，要证明这道题就需要作图。

表1.6　　学生证明图1,10问题的谈话记录

	S：我要证明∠A等于∠B。我想想看，你想要……？
*1	E：好……为什么你要画这条线？
	S：我还不知道。对，嗯……对，如果我画一条线……
	E：嗯。
*2	S：那就能平分∠ACB，这就能使我……得到两个相等的角……不。(停顿)
*3	对，这能使我得到两个相等的角，而且我得到了自反性，所以这个等于那个，好，我证明对了。
	E：先不要往下说，把所有这些都写下来，从你开始想画这条线时写起。
	S：好。
	E：我刚才问，你为什么要画这条线，你说你还不知道，你想想看，画出了平分线结果会怎样？
*4	S：对，我不得不试着这样做，我不得不使△ACB与△BCD全等。因为如果我能做到这点，那么∠A就等于∠B，因为全等三角形的对应部分是相等的。
	E：所以在△ABC中画了这条线，就是这个意思吧！
*5	S：不，还要得到一条公共边。
	E：对。
	S：还要得到两个相等的角。
	E：这就是为什么你要画这条线作为∠ACB的平分线了。
	S：对。

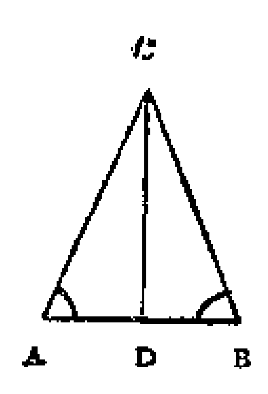


图1.10　给出表1.6谈话记录的那个学生所画的图

表1.6是一个学生证明这道题的谈话记录。这个学生开始总画一个三角形，三个顶点分别标上A、B、C，使$\overline{AC}$与$\overline{BC}$①相等，标出∠A和∠B，求证∠A等于∠B，正如图1.10所示。在谈话记录的*1处，这个学生从顶点C到底边引了一条垂线交点标为D。在*2处，这个学生推测$\overline{CD}$应是∠ACB的平分线，并且在*3处肯定了这一推测，这时他理解到$\overline{CD}$是两个三角形的公共边，这正是边-角-边式的全等三角形所需要的第二条相等的边。

这个谈话记录又提供了一个例证，说明可用不确定的目标结构证明三角形全等，而不是象前边所说的那样，使用确定的子目标，找出足够的全等成分模式。很可能是这样，如果这个学生有一个边-角-边的明确的子目标，那么只要画出$\overline{CD}$，它与其它成分的关系就清楚了。

进一步说，这个谈话记录还解释了作图这种相当不确定的过程。在*4和*5处表明学生已经打算要分成两个三角形，而求证相等的两个角分别在这两个三角形中，并且这个学生还认识到这两个三角形有一条公共边的价值。但是，我们且不说这种极为平常的打算，这个学生似乎对于求证中应该用添加辅助线的方法没有一个具体的计划。

① 原文为$\overline{AB}$，疑误。

表1.7　　学生解图1.11问题的谈话记录

	S：	对，我要在这儿画一条线。
	E：	好。
* 1	S：	这条线不太好，我应把这条线放在这儿，我应当放……画……或许、简直，我不能完全取决于这条线。看，我不能叫出这条线……我可以说这是中线。画出AX；这叫做中线，从A到CB的中线。
	E：	对。
* 2	S：	这正是中线的定义。好，你已知道，这正是一条辅助线。
	E：	对。
* 3	S：	我要记住已知条件，AB等于AC，这是已知条件。
	E：	嗯。
* 4	S：	我可以说AX等于AX。
	E：	嗯。
	S：	这是自反性。所以我可以说……我想想看，它等于自己，这个等于这个。
* 5		那么，我可以说因为AX是一条中线，它平分……在线段的中点相遇。所以我可以说CX等于BX。
* 6		这是定义……这是中线的定义。
	E：	了不起。
* 7	S：	那么我能说……△ACX 全等于△ABX……我可以说这是边-边-边式的。第6步就是∠C等于∠B……全等三角形的相应部分是相等的。
	E：	很好。现在我来问你一个小问题……你什么时候开始想在这儿画一条线。当时你想的是什么……。
* 8	S：	因为我知道，当我一看到这道题时……这是唯一的办法，因为已知条件中关于这两个角你再什么也没给，我只能用这种方法证明它们相等……如果这两个角是两个三角形的相应部分的话。
	E：	好。
* 9	S：	这是我能把这个三角形变成两个三角形的唯一办法……正好从中间分成两半。而且我还得到两个三角形的一条公共边。
	E：	很好。你什么时候想到使用两条底边的？

S：只要我画出这条线，第一件事……我写出已知条件后的第一件事……通常在证明中第一步要做的是……找出公共边。

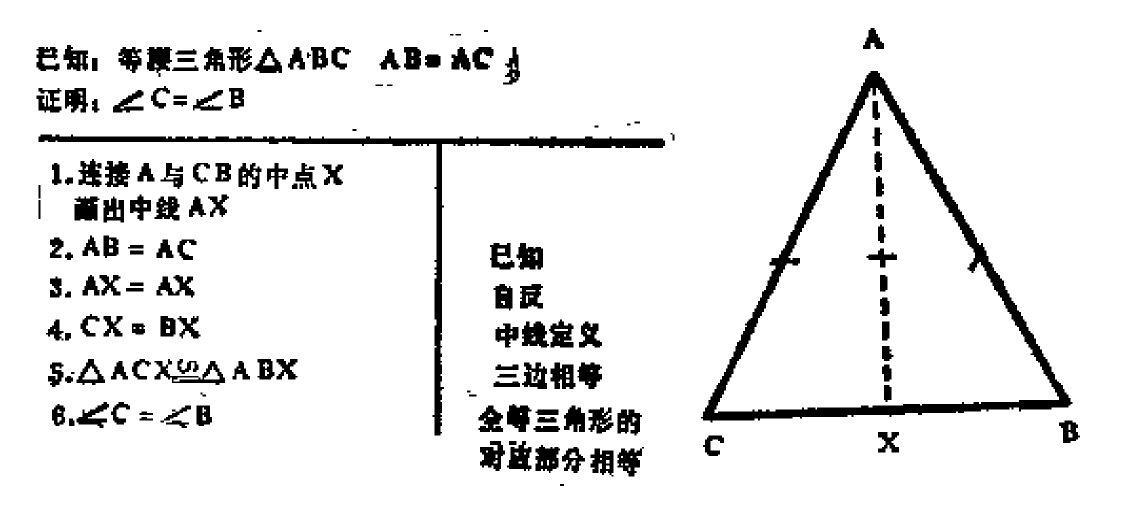


图1.11 给出表1.7谈话记录的那个学生所写出的证明和所画的图

表1.7和图1.11是另一个学生解这道题的记录材料——证明等腰三角形的两个底角相等。这个学生知道这是一条已经证明过的有关等腰三角形的定理，但是，我提出要学生自己证明。在教科书中所做的证明使用了角平分线而不是这个被试所使用的三角形的中线。因此，可以明显地看出，这个证明是通过问题解决得到的而不是凭记忆做出的。

这个学生先画了一个等腰三角形，三个顶点分别标上A、B、C，在图1.11的左上角画了一个表，把它的证明步骤写在此表中，标明$\overline{AB}$与$\overline{AC}$相等。然后这个学生画出线段$\overline{AX}$并且说明这是这个三角形的中线。在*1处表明这个学生还没有明确地想到引一条辅助线，完全是随意地说是一条中线。在*2处提到了"中线的定义"，这个学生写出这个定义然后又勾掉了。这一步大概与学生想到要遵守正式要求，每一步证明都要写出理由有关。在*3处，学生写出了第二步证明，在*4处学生写出了第三步证明，在*5处学生写出了第四步证明，在*6处写出了第四步证明的理由，在*7处这个学生完成了证明，记下了第五步、第六步。

在*8处和*9处表明，这个学生与前一个学生一样，在引辅助线时只有一个总的想法。这个学生显然想得到两个含有目标角的三角形，并且懂得这两个三角形应有一条公共边。还有一部分为证明三角形全等而寻找相应成分过程的记录未列出来，其中*10处表明对这条辅助线做不同的解释，就有不同形式的相等成分。在※11处学生对完成这个证明所需要的各种图形成分已经想出了一个具体计划的假设提供了一个十分有力的追述证据。

我们把如表1.6和表1.7所列的学生作业说明编成一套Perdix解题过程。在这个说明中有两个重要的概念。第一个概念是前面所说的逐级计划概念。第二个概念是由戈德斯坦(1974)提出的，涉及到使用表示类属特点的认知图式或把情境变成与图式一致的理想模式与过程。现在提出的作图理论使用了逐级计划的机制来决定作图的必要性，并且使用从动图解模式填充机制来选择正确的作图，以便以适当的方式改变情境。

戈德斯坦的MYCROFT程序可以解决儿童用LOGO教学语言书写乌龟程序时所出现的问题(佩珀特，1971)。乌龟程序的步骤是比较简单的，它控制着物体在一个空间移动。有一种程序设计，是使一种电动玩具的轮子可以按照已设计好的程序沿指定的路线移动。程序指令可以包括"FORWARD 50"，这就是指示乌龟移动50个单位的距离，或者包括"RIGHT 90"，这就是指示乌龟改变方向，向右转90°。还有一种程序设计，使用图形显示代替玩具乌龟。在这一设计中程序控制着指针在显示屏幕上移动，按指针移动的轨迹画线形成一个图形；例如，有一套指令可以画出一个木棍式的人形，或由一个代表树头的三角形和一条代表树干的垂直线组成的简单的树。

MYCROFT是一种在画图形的乌龟程序中表示纠正错误所需知识的程序。例如，可以编一个画树的程序，但是这个程序中可能

存在一些错误，如三角形的方位不对，或者没有树干与树顶相连，或者还有其它的错误等等。在MYCROFT程序中知识的表示包括几何图形的图解说明。每个图解包括这个图形各个部分所组成的表以及各个部分连结的方式。MYCROFT程序能把乌龟程序画出的图形与其原定图形的图解相比较。如果有不符合之处，MYCROFT程序有办法修正乌龟程序，使乌龟程序画出的图形与原定图形的图解特点一致。

表1.6和表1.7所示的作业可以用萨塞多蒂(1975)所设计的逐级计划过程来模拟，也可用戈德斯坦(1974)所设计的类似从动图解模式填充来模拟。首先，我们可以回想起来学生虽然加了辅助线但还不完全知道如何使用所加的辅助线。这说明在作图之前的计划过程是逐级的。在Perdix中，角相等目标首先要核实这两个角是否在两个三角形中，是否有公共点，是否由两条平行线和一条截断线构成的，或者是否允许建立链接角。那道证明底角定理问题的几何图形中这些先决条件没有一个得到满足，这样一来Perdix就没有一项计划可采纳了。因此，当不必要作图时Perdix和人类问题解决者似乎都以选择一个总计划这同一方式来做出决定，把已知的三角形变成两个三角形，而不是平行线或其它某种作图。一般过程可能是如下三步：

1. 检验可行计划的先决条件，哪一个计划的先决条件得到满足，就选用哪个计划。
2. 检验通过作图满足计划先决条件的特点。选择一套能被满足的先决条件。
3. 执行那个先决条件得到满足的计划。

表1.6和表1.7的记录对这种问题来说是具有代表性的并且表明只有做完了第二步后，才能详细地做第三步。这与分级计划的概念是一致的。

第二，戈德斯坦提出的从动图解模式填充概念对于学生进行必要的作图提出了一个似有道理的假设。具有一条公共边的两个三角形的图解应包括两个三角形、五条边、四个点做为主要部分。如果我们把上述部分分别标上T_1、T_2、S_1、S_2、S_3、S_4、S_5、P_1、P_2、P_3、P_4，那么这个模式所需要的关系就应成对表示，指明哪个点是哪条线段的端点〔(P_1、S1)(P_1、S_2)(P_2、S_2)(P_2、S_3)(P_2、S_5)(P_3、S_3)(P_3、S_4)(P_4、S_1)(P_4、S_4)(P_4、S_5)〕并指明哪一线段是哪个三角形的哪一条边〔(S_1、T_1)(S_2、T_1)(S_5、T_1)(S_3、T_2)(S_4、T_2)(S_5、T_2)〕。

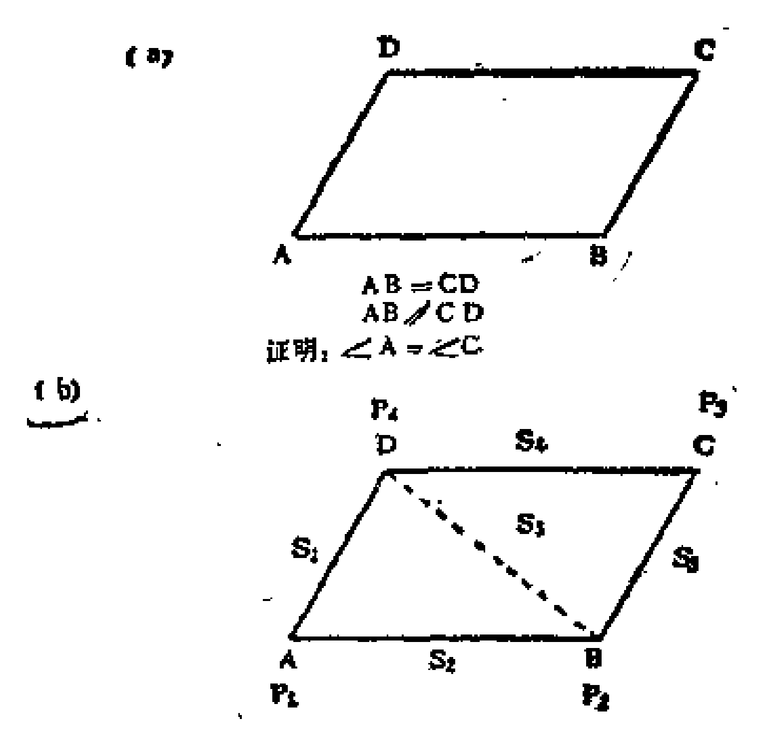


图1.12　需要添加辅助线的简单问题的图形，其各部分与正文中所述的图式相对应

我们来看图1.12(a)格中的问题。要证明这道题就需要引一条辅助线；所需要的这条辅助线是对角线$\overline{BD}$。学生很容易地做出了这道题，稍加思考后很快就画出这条对角线。表1.8列出学生解这道题的谈话记录，在* 1 处引了辅助线，从后边的过程来看，显然在详细证明之前要先引辅助线。假定这个学生有上述所说的图解，

并且选用的计划是通过证明三角形全等来证明其所含的两个角相等。这就需要有两个三角形，并且每个三角形含有一个所要求证的角。这就是说一个三角形有$\overline{AB}$和$\overline{AD}$两条边，另一个三角形有

表1.8　　学生解图1.12问题的谈话记录

E：现在我让你证明一个定理，你不用列出所有的证明步骤。好好想想，看怎么做。如果四边形的两条边相等且平行，那么这个四边形中相对的两个角相等。

对。

E：想想看，你能做吗？怎么做？

S：好，你有一个四边形……

E：对。

S：并且……对，已知有两条边平行且相等

E：嗯。

S：要我证明相对的两个角相等？

E：嗯。

*1　S：所以我引了一条线……

E：对。

S：我要说明这个角……不，我要证明这个角，这个角等于那个角。

E：好。

S：对，因此我……(停顿)……这两条平行线……

E：嗯。

S：对，这两条平行线是已知的。这是四边形的定义。还有这两个角，我要证明这个定理，也就是说两个对角的定理。

E：对。

S：对，我要……我要标出这两个角。

E：嗯。

S：一个，两个。

E：嗯。

S：我通过……证明这两个角相等。

E：对。

S：……内错角。

E：对。
S：嗯，我不用把全部过程写出来。（停顿）
E：对，很了不起。
S：那么我就用边-角-边式的方式证明。
E：对，这很好。
S：那么我就得到相对应的部分。
E：对。

$\overline{BC}$ 和 $\overline{CD}$二条边。这种情境就把这个图解的部分与下面的映射匹配起来：P_1是A，P_2是B，P_3是C，P_4是D，S_1是 $\overline{AD}$，S_2是 $\overline{AB}$，S_3是 $\overline{BC}$ *，S_4是 $\overline{CD}$。我们看到S_1、S_2、S_3和S_4之间所需要的连接得到了满足。情境中缺乏与S_5相应的其端点应是B和D的成分。显然要补上所缺的成分，就要引一条辅助线 $\overline{BD}$。正如图1.12(b)格中所表示的那样。

在证明等腰三角形二底角相等的问题中，按刚刚说的把图形映射成图解可能稍复杂一点。最初图形中只有用线段相互连接起来的三个点，而不是四个点。这就意味着Perdix一定要能把这个三角形的底边分成两半，以便作为所需要的两个三角形的底。对人类被试者来说这样做显然困难不大或根本没有困难。很有趣的是，在我们的研究中没从一份学生的谈话记录中发现一般认为是比较漂亮的证明，那就是需证明全等的三角形是△ABC和△BAC。我推测对学生来说采用把一条线段分成两半这种办法，可能比从不同角度来识别同一个三角形更容易。在任何情况下，如果我们采用了图1.1所示的标号，那么这个图形就可映射成如下图解，C为P_1，A为P_2，B为P_3，$\overline{AC}$为S_2，$\overline{AB}$为S_3。附加点X提供了P_4，以及S_1（$\overline{CX}$），S_4（$\overline{XB}$）。完全象在四边形的问题中一样，必须要有S_5（$\overline{AX}$）。

* 原为$\overline{BE}$，疑误。

我们已做过两个实验，都是让学生解需要简单作图的几何题。实验结果与我就问题解决中可以作图这一方面所提出的看法基本一致。一个实验表明，当问题的情境特点和被试认知方式的特点之间有比较直接的关系时，作图是比较容易的，而当问题的情境特点与被试的认知方式之间的关系比较复杂和曲折时作图就不那么容易了。第二个实验表明，作图的过程由于沿用同一模式总计划的最新使用而受到促进，因为这可刺激作图。

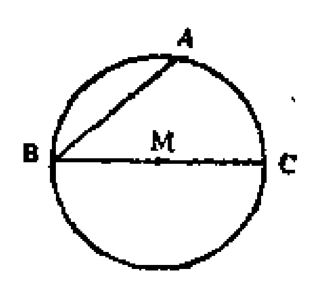


图1.13　关于添加辅助线实验所用问题的图形

马冈和我做过一个实验，比较了两个问题的难度。两个问题都使用了图1.13中的图形。M点是圆心，$\overline{BC}$是直径。给第一组被试的问题是，已知$\overset{\frown}{AC}$是80°，求∠ABC的度数。给第二组被试的问题是，已知∠ABC是40°，求$\overset{\frown}{AC}$的度数。要做这两道题都要引辅助线$\overline{AM}$。这就形成了圆心角，圆心角的度数与其所截取的弧的度数相等。根据一条弧的度数的定义表明圆心角和它所对应的弧相等。引$\overline{AM}$这条辅助线还形成了一个等腰三角形△MAB，其两个相等的底角之和等于M形成的外角。

这个实验的立意依赖于这样一个事实，大多数学生解题的思路是正向的，也就是从已知条件开始，而不是逆向的，从目标开始。这就意味着，当$\overset{\frown}{AC}$的度数已知时，大多数学生设法在图形中找出与这条弧有关的东西；而当∠ABC的度数已知时，大多数学生设法在图形中找出与这个角有关系的东西。我们推想已知弧的信息比已知角的信息可能更有助于引出$\overline{AM}$这条辅助线。这是因为这条辅

助线是圆心角模式的组成部分，并且必定与弧的度数有关。而要把角与作图联系起来，就要想到一个等腰三角形，但是∠ABC与等腰三角形模式之间的联系似乎不如$\overset{\frown}{AC}$与圆心角模式之间的联系那么直接。

在我们的实验中，把大学生被试分成六组。首先让学生复习一下解如下无论哪一类问题所需要的几何概念。一类问题是证明等腰三角形的二底角相等，或证明三角形的一个外角和其它二内角之间的关系，另一类问题是证明圆心角的度数与它所对应的弧的度数相等。看图1.13，或者弧的度数已知，或者圆心角的度数已知，而其它的量要去找出。结果表明，拿到弧的度数已知这道题的52个被试中有73%的人做出了这个题，反之拿到角的度数已知这个问题的53个被试中只有47%的人做出了这道题。这个差异是可靠的；这两个比例之间差异的95%的置信区间是0.26±0.19。结果证实了我们的预想，角→弧的形式比弧→角的形式更难。如果我们接受了关于这二种信息与有关的认知图式之间的关系的分析，那么结果就为支持如下观点提供了证据：作图是由把问题的特点与作为模式填充基础的图解联系起来的过程产生的。

蔡科林(Seth Chaiklin)和我所做的另一个实验是关于问题解决定势对需要作图问题的影响。对作图所进行的理论分析，在很大程度上依赖于在这个系统中执行计划的过程。假定在计划过程中看出了需要作图，并且用代表计划过程特点的那种自顶向下的形式来作出作图的决定，那么可以料想到，需要作图的问题也象在其它问题中所表现的那样同样易受定势的影响，制定计划时倾向于接受最新体验的影响。

蔡科林和我安排好了两套归纳问题。其中一套问题要用由平行线所形成的各种角的关系来解。另一套问题要用全等三角形来解。在这两套归纳问题之后，有一道需要作图的问题，而这道可用

两种不同的方法解。作图的一种结果是形成两个可证明为全等的三角形，另一种结果是形成平行线中的各种角，并用这些角作为链接角来解这道题。显然可以假设，那些做过归纳平行线问题的被试将通过作图形成由平行线联系起来的链接角，而那些做过归纳全等三角形问题的被试则通过作图来形成两个可证明为全等的三角形。这些被试做完了这道可用不同方法解答的问题之后，再让每组被试解答对偶组做过的最后一道归纳题。每组被试要做这道题就需要使用不同于他们已经用过的问题解决计划。这两道题都不需要作图，这就使我们有机会进一步看到在一般计划水平上问题解决定势的倾向性效果。

被试都是大学生，我们给每个被试几道复习题，使他们恢复一下证明三角形全等所涉及的概念以及证明两条平行线为一条直线所截而形成的角的定理。在入门问题之后，两组被试各接受一套归纳问题，每套都是14道题。六个被试各自分别做这两套归纳题中的一套。

每组的第十五题如图1.12所示，此外还已知 $\overline{AD}$和 $\overline{BC}$相等且平行。解这道题有两种很容易的方法。可引对角线 $\overline{BD}$，结果根据边-边-边的模式$\triangle ABD$和$\triangle CDB$全等，或者延长 $\overline{AB}$，$\angle A$就等于这条延长线所形成的外角，因为这个外角与$\angle A$是同位角，而这个外角又等于$\angle C$，因为它与$\angle C$是内错角。

正如前边所说的，每组还有一道附加题，这道附加题不能用他们以前用过的一般计划来解。把图1.14(a)的那道附加题作为第十六题给那组做过归纳三角形全等问题的被试。使用平行线的有关定理很容易就解出了这道题。$\overline{AB}$平行于 $\overline{CE}$，因此，$\angle CAB$与$\angle ACP$互补。把图1.14(b)的那道附加题做为第十六题给那组做过归纳平行线问题的被试。这两个三角形由于有对顶角的关系所以以边-角-边的模式全等。现在对两组被试来说，各自所接到的第十六题

就是对偶组所做过的第14项。

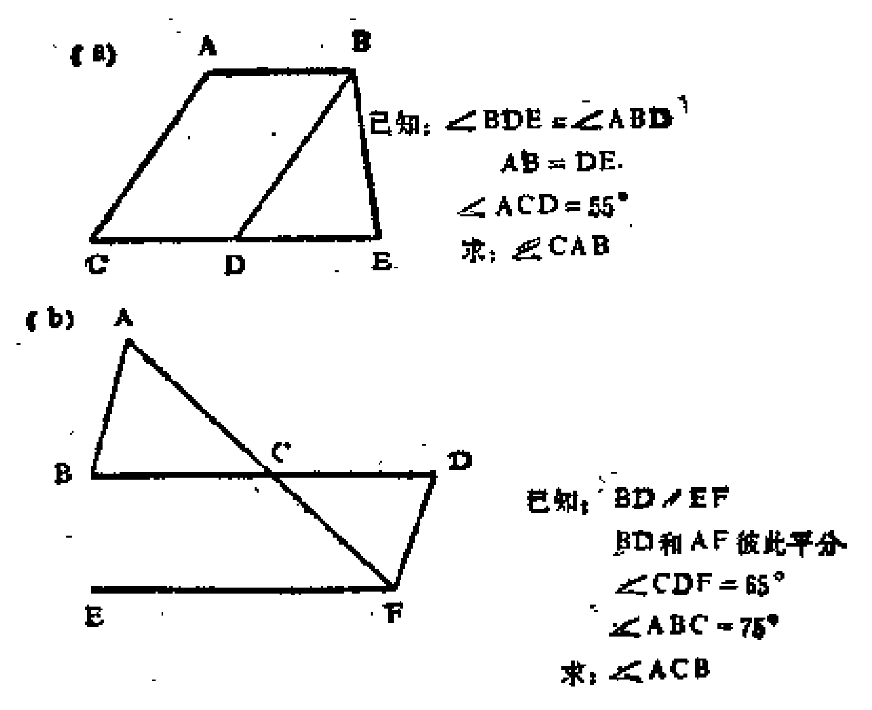


图1.14 关于添加辅助线和定势实验所用的问题

这种实验控制产生了足以可信的差异，即使每种实验条件下只做了六个被试。在证明平行四边形的对角相等的第十五题中，做过归纳三角形全等问题的那组被试，六个人中有五个人是通过引对角线并证明所形成的角相等而解这道题的。而做过归纳平行线问题的那组被试，六个人中只有一个人是用了这种方法解这道题的；其他五个人都是用平行线所形成的角的关系解了这道题。(符号检定其显著性 $P<0.02$，这里被试的解和归纳顺序之间的随机事件是一致的。)

图1.14(b)的问题必须通过证明三角形全等才能解。做过归纳三角形全等问题的那个组，六个被试一上来就全都采用了证明三角形全等这种解题计划。而在做过归纳平行线问题的那组被试中，只有二人一上来就采用了证明三角形全等的解题计划；其余四人设法使用平行线的关系来解题。(这也是根据符号检定来看其显著性的，$P<0.02$。)

图1.14(a)的问题可用平行线的有关定理来解。在做过归纳平行线问题的那组被试中，六个人中有四人一开始就使用引平行线的解题计划，还有两个人没有引出这条平行线，但是提到可把引平行线作为一个子目标。在做过归纳三角形全等问题的那组被试中，六个人中有三个一开始就使用了证明三角形全等的解题计划，只有一个人使用引平行线的解题计划，还有两个人一开始就使用了四边形内角之和的解题计划。(我们看到这些被试，他们一开始或者采取证明三角形全等的解题计划，或者采取引平行线的解题计划，从符号检定来看二种条件之间的差异是显著的，$P<0.04$。)

这个关于定势的实验为作图的过程和制定计划的过程之间有极密切的关系这一看法提供了一定的支持，除此而外，还进一步说明了问题解决定势对制定计划过程的影响。关于作图问题的研究还在进展中，关于作图过程的模拟模型将被加到现有的Perdix程序中，目前正在做进一步的实验工作。因此，目前的研究似乎是很有希望的，可望根据逐级计划和图解基模式填充的概念对几何问题中进行简单作图的过程做出比较合理的解释。

Ⅶ. 该研究对教学的意义

我在这篇文章中所报告的研究是描述性的和分析性的，而不是规定性的和发展性的。我企图了解和表示学生在学习几何课时所需要的知识，而不是决定应该教什么或怎样教。研究的结果不能为推荐某种教学实施提供基础。但是，可以通过一些研究结果来讨论几何教学内容中的一些问题。

我所讨论的解几何题所需要的三种知识就关系到几何教材编写和课堂教学的内容问题。我讨论的主要之点是，应当看到关于模式识别的知识和关于定理推理的知识我们是明确地教给了学生

的，而关于策略知识没有明确教给学生。我讨论了策略知识的一般特点，并且把第Ⅵ部分所谈到的各种特定内容方面的策略与通常所考虑的一般问题解决策略区分开。最后我要谈谈在几何课中越来越多地把问题解决的策略明确地教给学生，可能产生的利和弊。

首先，看看知觉模式识别所需要的知识。关于这种技能的教学主要是通过各种能锻炼学生找出需要识别的视觉特点和关系的图形和练习来进行的。例如，用图1.15这种图形使学生掌握二条平行线为一条直线所截形成了一些有关的角。通过这个图形让学生识别出每对对顶角、同位角、内错角、同旁内角。在大多数教科书中，关于这种模式识别的教学很少使用言语描述，而是用实例来说明各种视觉模式，学生通过选择合适的概念实例掌握了有关的特点。

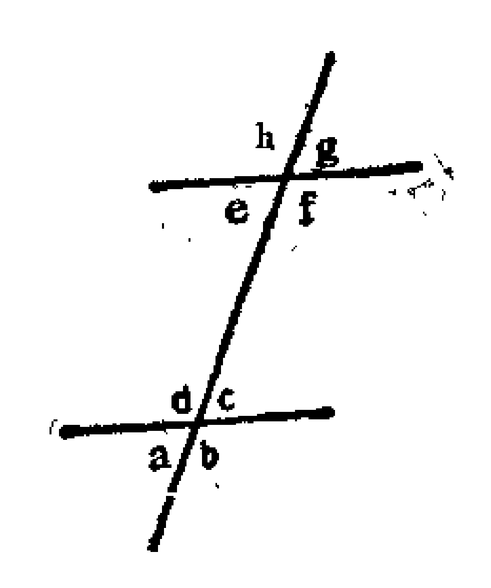


图1.15 指导模式识别所用的图形

问题解决过程中的推理所使用的定理也是通过明确的教学传授给学生的。象"对顶角相等"和"三角形内角之和等于180°"这样一些定理或以公理或定理的形式提出，并以明确的方式加以叙述。

在我调查过的所有教科书中都没有明确地提出关于提出目标

和制定计划所用的原则。我猜想，一般认为象表1.2中所列举的提出目标和制定计划的规则不应是几何课教学的内容，对此我表示怀疑。许多人或许认为，几何这门学科的内容仅是由我已描绘为模式识别和推理知识的概念和定理组成的，而我所说的策略知识则构成了学生会运用所学过的知识的能力。还有一种观点认为概念、定理和策略都是学生在几何课上应获得的知识，有些学生对学几何课感到困难可能是因为他们没有获得所需要的策略知识的缘故。

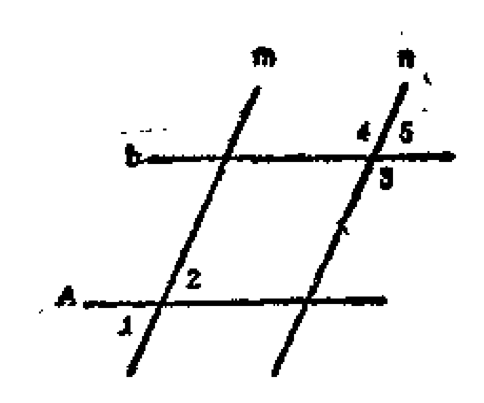


图1.16 给出表1.9谈话记录的那个学生所画的图

表1.9 学生解图1.16问题的谈话记录

S：好，已知a和b平行于m和n。
E：不对，那么……
S：A平行b，m平行于n。
E：对。
S：P等于40°，求q的度数。嗯，我应该给这些角起个名。
E：对，可能最好你先画一张小草图，这样你才能给你需要的角起名。
S：对，我要画，而且，P要标上1。
E：对。
S：那么，2。
E：把2放在那儿，对。
S：那么3。不，等一等，3和4，我猜猜。

E：对，为什么你把2'放在那儿？

S：嗯……我不知道。用得着对顶角。

E：对。

S：等一等，你不是要我写出证明吗？

E：不，你可以写出你想写的一切。而我主要关心的是你思考问题的办法。

*1 S：好。我想我要做的第一件事是设法……我要设法找出有否内错角或同位角？

E：对。

S：或者其中一种。

E：嗯。

S：我猜我应当说……对，等一会儿。我猜我大概应该在那标上5。

E：对。

S：我不知道我要否这样做。

E：可以。

*2 S：这两个角互补。

E：对。

*3 S：这没有多大帮助。那么，∠5的度数……它等于∠1的度数吗？

E：好，你一定能做出来。

S：怎么……如果这等于……这等于40°。

E：对。

*4 S：好，等一等，度数……我不能，我不知道。我不知道怎么做。

E：好。

表1.9的谈话记录就是一个因缺乏策略知识所致结果的例子。图1.16是这个学生自己画的。问题还是图1.1所示的那道题，本文开头所举六个被试的正确答案已在表1.4中列出。而与这个学生就上述这道题交谈的过程中，我还问了许多关于两条平行线为一条直线所截形成的角之间的关系问题。这个学生对我的提问所涉及到的所有成对关系的知识的掌握是令人满意的。我推想这个学生所以没有做出这道题是因为缺乏构成具有相等关系的链接角的策略，这就是表1.2中的第4条规则所示的策略。在表1.9的*1处似乎

表明这个学生掌握了第2条规则所示的策略。企图在可推出有定量关系的角之间找出几何关系。在* 2 和* 3 处似乎表明这个学生掌握了第2条规则所示策略的一部分，也就是说如果两个角都与第三个角有某种定量关系就可以推断出这两个角之间也具有定量关系。但是，在* 4 处，学生中止了解题。第4条规则所示的策略是通过与∠P的已知关系找出某个与其相等的角。这个学生做不出这道题显然是他没有掌握第4条规则的策略知识。

问题解决所需要的策略知识是可以学习的。刚刚说过的那个未做出图1.16那道题的学生在二周以后，就能做出另一道也需要找出相等的链接角的问题。在理论上和实践上所关心的一个问题是在没有明确地教学生策略知识的情况下学生是怎么获得策略知识的。一个合理的推测是学生能通过对教科书中的例题和教师的例证进行归纳而学到策略知识。有些学生也可能不用通过对含有必要规则的例题的归纳，自己就会得出问题解决所需要的策略，但是多数学生似乎要依靠例题去获得必要的策略知识。虽然关于提出目标和制定计划的原则没有明确地包括在几何课的内容中，但是它们却以所给例题的性质这种隐含的形式出现了。

对某些读者来说把策略与概念和定理一样看作是一种知识形式似乎不太好理解。策略知识是一种技能形式，所以策略如同知道如何做饭或知道如何行船一样同属一种知识范畴。把熟练作业的知识表示为一种过程这是很自然的，并且常常把它看作是不同于概念知识和定理知识的。但是，正如威诺加德(Winograd)(1972)所指出的，概念知识和定理知识也自然能被表示为各种过程，而过程性的表示在问题解决中是最为合适的。确实在Perdix中，是用过程的形式来表示概念和定理的。概念被表示为模式识别的过程，定理被表示为进行推理的产生式规则。而策略知识基本上不同于上述二种知识，那二种知识普遍被认为是几何课教学的内容。

解几何题的Perdix中所表示的策略知识比起通常所说的问题解决策略更特殊。例如波莱Polya(1957)找出了一些一般的策略过程，如找出问题的未知的和已知的数据，如果眼前的问题没有进展时想想一个比较简单的有关问题。另一个例子是兰达Landa(1976)所做的推荐，学生应当想想一个问题的各种成分彼此联系的各种方式。在此处的分析中找出的策略是比较特殊的，利用了几何这个问题领域的特定内容。

明确地教特定内容方面的问题解决策略可以是符合意愿的或者是不符合意愿的，但是应当假定是能教的。解几何题所用的策略原则并没有什么神秘。所需策略的一种形式就是以本文所述的问题解决模型的明确方式叙述的。只要能给策略规则一个明确的叙述，那么编写出能把这些规则传授给学生的教材，并让学生练习使用这些规则就成为完全可行的了。

进行特定领域的问题解决策略的明确教学其主要优点是会使那些在目前教学情况下感到困难的学生的作业得到改进。一种可能的好处是对相当数量学不好几何课或由于学不好几何课而妨碍了其它数学课学习的学生，施以解几何题所需策略的教学，会得到满意的效果。

进行问题解决策略明确教学的第二个优点是可以使学生，或许也可使教师更好地鉴别问题情境的性质。策略原则表明了问题解决步骤的目的，因此对问题求解各个部分彼此之间的联系方式和与主要问题解决目标之间的联系方式提供了一个更好的理解。

进行问题解决策略明确教学的一个主要缺点可能是减慢了教学的进度。还不清楚现在这种解例题和以隐含的方式传授策略原则所花的时间对进行策略原则明确教学是否够用。如果每项几何内容的教学所需的时间确实因进行策略教学而增加了，那么就需要权衡一下，对问题求解结构的进一步了解，其所获是否足够弥

补因进行策略知识明确教学而造成的几何内容教学的减少。

进行策略知识明确教学的第二个缺点可能使学生养成过于机械的问题解决作风，特别对那些自己能发现策略原则或能从例题中归纳出策略原则的学生来说更是如此。从目前所进行的教学来看，所有学生如果最终都能掌握了问题解决策略，那么他们完全是通过发现法而学到的。给学生提供发现策略原则的机会是否会成为有助于提高学生使用这些原则并把这些原则迁移到新情境中去的能力的一种重要途径，现在还不清楚。但是这只是一种可能性。对那些解了一系列问题但自己还归纳不出这类问题所需策略的学生，分别采取矫正教学的形式，或许是在课堂上进行策略教学的理想方式。

Ⅷ．对一般问题解决理论的意义

我研究几何问题解决的主要目的是想获得关于问题解决过程的新看法。我希望能对问题解决过程有一个更好的了解。显然目前已有了进展，导致了三个方面的理论发展。一般问题解决理论的第一个争论点是良结构和不良结构问题之间的区别。似乎是通过对不确定目标结构和需作图的几何问题解决过程的分析，更加清楚地了解了不良结构问题解决的过程。一般理论所关心的第二个争论点是问题解决的理解过程，涉及到完全正确地解题与理解情境和目标之间的差别。把问题解决表示成一个模式匹配和模式生成的过程似乎为理解过程若干方面的分析提供了一个框架。最后，几何为研究知觉和思维之间的关系提供了一个特别合适的领域。在几何问题解决中，知觉过程的重要性是很明显的，因为问题的提出需要使用几何图形。作为问题解决过程组成部分的知觉过程的作用正是通过对计划过程的分析而开始得以阐明的，包括计划

选择过程中关键的全局特点的使用和图形成分检验过程中知觉扫描过程的使用。

A.良结构和不良结构的问题

大多数问题解决的实验研究使用了象梵塔(塞蒙,1975),水缸问题(阿特伍德、波尔森,1976),和运输问题(格里诺,1974;塞蒙、里德,1976;托马斯,1974)这样一些相对简单的问题。这些问题都是良结构的,有一个已知的初始情境是确定的,一个将要达到的目标情境也是确定的。能用来改变情境的操作已被规定,初始问题情境包括了达到目标所需要的所有成分。

但是,还有许多重要问题不属于这种良结构类型的问题。我们想知道通过良结构问题的研究所得出的结论是否适用于不良结构的问题,目前已对不良结构问题解决做了一些研究。里特曼(1965)分析了一个作曲家写赋格曲的大声思维记录。还对国际象棋手(蔡斯、塞蒙,1973;德格鲁特,1966;纽厄尔、塞蒙,1972)和围棋手(里特曼,1976;里特曼、威尔科克斯,1976)做过一些谈话记录分析和实验研究。但是,目前研究不良结构问题解决所得出的分析还远不如研究良结构问题所得出的分析那么完整,并且在良结构问题解决中起作用的原则是否在不良结构问题解决中也有效,是否不良结构问题的理解需要根本不同的原则,这是一个未解决的问题。

几何问题是良结构的,与作曲和下棋这些问题比较起来尤其如此。一般说来,问题的已知信息提供了明确的初始情境,对于所求证的或所要找出的未知量给予了明确的叙述,并且可使用的操作与以前在课堂中学过并证明过的命题和定理相一致。但是,通过对几何问题解决中两个重要特点的分析似乎为次良结构问题解决中可能出现的各种过程提供了信息。进一步说,在几何问题

中这些过程出现的形式是比较简单的，因此对这些过程的性质似乎特别容易理解。

不确定的目标结构和作图这些几何问题的特点似乎对解不良结构的问题有启发，因这与不良结构问题一般所具有的两个特点有关。不良结构问题的目标常常是不明确的。例如一个棋手可以决定向对手的皇后发起进攻。一个作曲家只根据作品的形式、演奏的乐器以及所构想的主题开始作曲。这些都是高度不确定目标的例子。相比之下，在良结构问题中，问题解决者能把眼前的情境与每一进展阶段的目标进行比较，而象"进攻对手皇后"或作"赋格曲"这种不确定目标就没有提供一个能用来作为问题解决活动目标的特定情境。

可以回想一下，学生正是使用了一个不确定目标表示解答了需证明三角形全等的几何题。而具有以模式识别系统的形式来表示知识结构的Perdix就模拟了学生的这种作业。似乎理所当然地推断出，在不确定目标结构方面熟练的问题解决者可能具有复杂的模式识别系统，使得他们有可能对问题解决中所出现的复杂情境进行判断。有确凿的证据表明，模式识别是下棋技巧的一个重要因素。德格鲁特(1966)，蔡斯和塞蒙(1973)曾表明，让象棋大师与虽参加过比赛，但水平低于大师的棋手们看五分钟棋盘，然后回忆棋盘上的局势，结果前者优于后者。根据这种作业可推断出，象棋大师对识别棋子的模式有超常的知识(塞蒙、吉尔马廷，1973)。目前，还没有一种理论分析能说明象棋大师在对奕时所进行的问题解决是如何使用了渊博的模式识别知识的。但是，似乎可以理所当然地推断出，他们正是用模式识别来评价棋盘局势的结构特点，而我们也正是用模式识别来发现每位棋手的潜在优点的。派特曼和威尔科克斯(1976)通过对围棋名手的初步分析强调指出，模式识别是下棋中影响棋手提出目标的一个重要因素。虽然用来表示

证明三角形全等这个目标的模式识别系统在复杂的问题解决模式识别中是一个相对简单的实例，但是，对良结构问题中表示目标的模式识别系统进行理论概括似乎会对许多不良结构的问题解决提供一个重要的原则和启示。

许多不良结构问题除了有目标不确定这个特点外，第二个特点就是问题的已知条件不全。例如，作曲家在开始作曲时只对乐曲的几个部分，或许包括决定试用的主题思想有个总的构想。大部分作曲都是在曲调发展到那一步时才随时加上新的音乐素材的。有一作者曾经记述了作曲家作曲时的情况"他们只怀有一个音乐思想并面临着如何展开的问题，关于'展开'这个问题正是作曲的核心"(亚伯拉罕，1949，PP.4—5)。相反良结构问题在初始情境中就给了一组目标，通过重新安排这些目标而改变情境的规则和由已知目标的某种特定安排而构成目标情境。作曲也象写文章和搞设计一样，问题解决者必须获得相当多的答案所必需的材料，并且要决定如何安排这些材料。

需要作图的几何题为情境不全问题的作业提供了一些新的信息。在这些情境不全的问题中，问题解决者必须补充上问题的初始情境没给的那些成分。从分析Perdix的解题过程看来，正是使用了逐级计划的思想，包括为使计划可行所要满足的先决条件的知识，找出当前情境与必要的先决条件之间的差别的产生式，生成所需成分的过程。制定计划和模式填充的概念似乎与里特曼(1965)研究一个作曲家时所看到的问题解决过程的许多特点是一致的。至于不确定目标结构，我需要对一些不良结构问题的作业进行透彻分析才能确定作图与更加复杂情况下的创造性过程之间的类似程度。但是复杂的创造活动所需要的基本过程与以简单几何题形式所表现的过程好象类似。

关于良结构和不良结构问题的一般争论，几何题的分析似乎

支持了这样一个结论：问题的复杂性将随着结构良好性的程度而有所不同。几何题比起简单的难题来似乎需要更复杂的机制，但是说明几何问题解决所需要的理论程度是十分有节制的。进一步说，似乎理所当然地推断出，解几何题所用的模式识别目标结构。逐级计划和图解基模式填充这些机制，或许是解决更加复杂的不良结构问题所用的各种机制的简单形式。因此，如果解不良结构问题的过程与解简单的良结构问题的过程相比只是更复杂的话，那么这两种结构的问题解决原则似乎就没有根本的差别了。

B.问题解决中的理解过程

在目前有关问题解决的讨论中理解过程不是一个主要的议题，不过最近有些研究对基于原文表象的表示问题的过程提供了新的重要分析(海斯、塞蒙，1974，1977；诺瓦克，1976；塞蒙、海斯，1976)。但是，格式塔心理学家关于问题解决的早期著作强调了理解的重要性。魏特海默(1959)特别强调机械地使用算法解题与理解情境结构的解题之间的差别。

象沃特海默这样的格式塔心理学家所提出的看法是很有说服力的，但也是很笼统的。通过理解来解决问题，从目前的研究看来在很大程度上还是一种轶事，还没有得到导致理解的认知过程分析的支持。本文所报告的关于几何问题解决研究的目的之一就是提出一个有关理解过程的清楚的说明。最近通过对语言理解的分析提出了语言理解系统，这些语言理解系统对语句、段落或故事形成了综合表示。在这些语言理解系统中，把理解隐约地定义为领悟信息各成分之间的关系并对这些关系构成表示的过程。

由一个语言理解系统所构成的这种表示如图1.17所示的例子。这个表示是由诺曼、拉梅尔哈特和LNR研究组(1975)开发的计算机系统构成的。图中所表示这个句子是"Ida borrowed a tab-

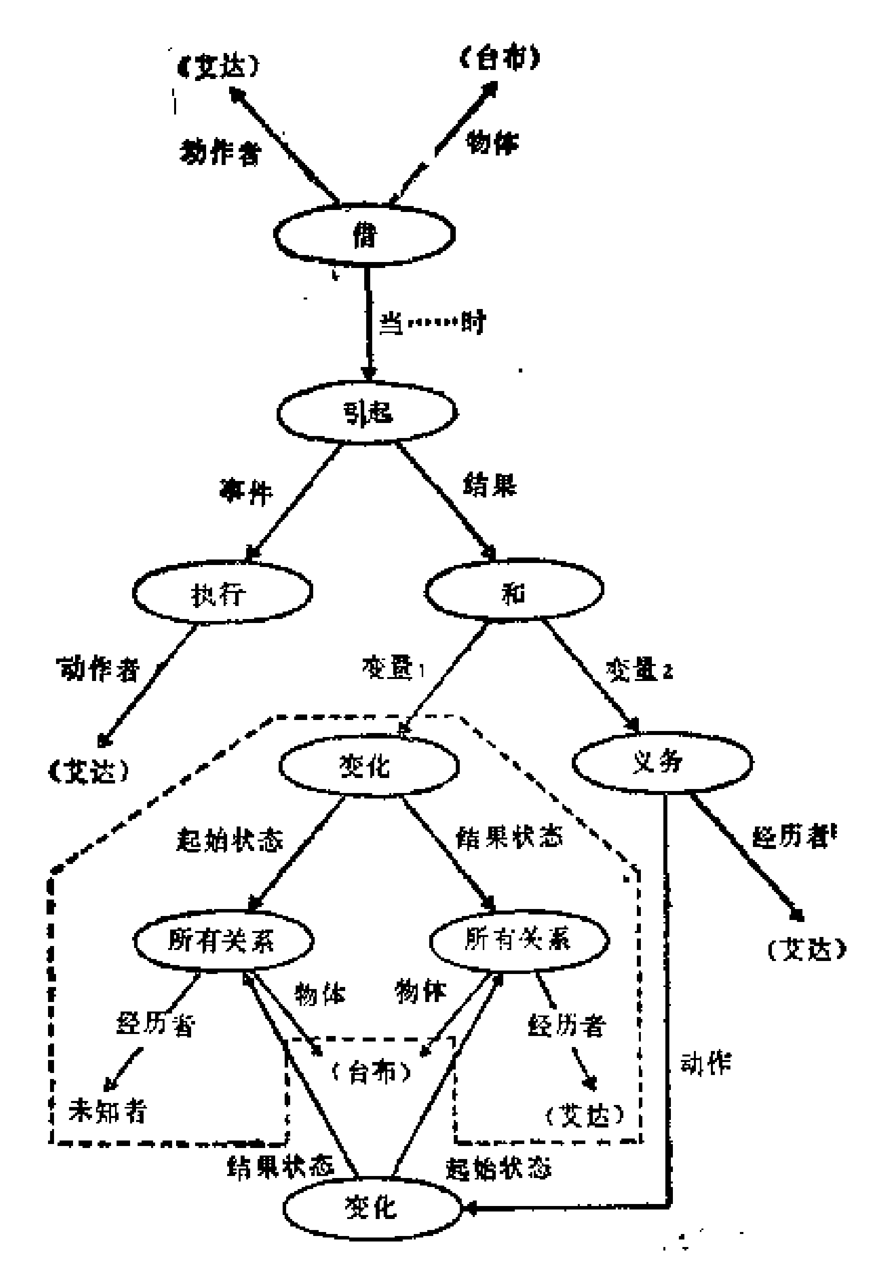


图1.17 用一个语言理解系统构成的"艾达借一块台布"的表示

lecloth"(艾达借了一块台布)，在这个句子中所包含的各种特定成分和其它涉及所有关系动词的句子都是由金特纳(1975)提出来的。该系统把"Ida borrowed a tablecloth"这个字符串作为输入接受进来，然后执行一个语法分析过程，结果找出"Ida"(艾达)是句子的主语，"borrow"(借)是谓语，"tablecloth"(台布)是宾语。"borrow"的图解表示被存贮在系统中，根据这个图解系统构成的表示表明Ida是动作者，tablecloth是被借的物体。除了Ida、borrowing

和tablecloth之间的这些基本关系之外，理解句子的意义还包括进一步的信息表示。这种表示包括Ida执行了一个动作引起了tablecloth所有关系变化的信息，Ida借到了tablecloth后就成了具有tablecloth的人，Ida还有义务改变tablecloth的所有关系，使tablecloth回到原主那里。

我们可以讨论一下图1.17所示的这一特定内容的表示，但是图中所示的这种特定的关系结构不如语言理解理论的某些一般特点重要。语言理解研究者最近发现了一套有关构成一般特点表示所需过程的一般原则。在诺曼和拉梅尔哈特的系统中，所使用的主要知识是动词意义的图解表示，因为这个系统的运行是靠给作为表示句中所述主要动作和关系的动词变量的各种概念赋值而实现的。

本文所报告的问题解决模型提供了一个在问题解决中关于理解的表示，这种表示类似于诺曼和拉梅尔哈特的系统中关于语言理解的表示。这是因为在解题过程中，Perdix所用的各种成分之间构成了一种关系模式。回忆一下图1.1，在该图的(h)格中就表明了问题被解决时所构成的表示。这种表示类似于图1.17的形式，虽然这不很重要，因为二种表示都只是具有标号弧线的网络，是一种很一般的表示形式。重要的事实是图1.1的表示是作为解题过程的部分而产生的，因此把理解过程表示为问题解决过程的一个部分。Perdix生成的结构类似于语言理解系统构成的结构这决不是偶然的。编写Perdix程序的框架就是ACT程序系统的一种变式，ACT程序系统是由安德森(1976)为了分析语言加工和语言获得而开发的。

Perdix生成的问题表示包括问题情境各成分之间的适当关系。对一个问题解决者来说常常提出的另一个问题是在解题过程中是否理解了该做什么。这个问题涉及到问题解决过程中所执行

的各动作之间的关系，而不是问题情境的各成分之间的关系。

可从两个方面来考虑一套动作的理解。一个方面是考虑动作者执行各种动作成分所要达到的目标。这可以称之为理解动作的功能方式或有意方式，因为这个动作的表示指明了每个动作成分是为了什么目的，或者动作者执行每个动作成分企图达到什么目的。第二个方面是表示对每一步动作进行证实。这可以叫做理解动作的说明方式，因为这种表示指明了每一个动作成分为什么要完成，同时指明了这个动作怎样进行。

图1.1所示的表示对这道题的解题动作提供了部分的说明性理解。每个重要的步骤或者是推出角与角之间的定量关系或者是推出一个角的度数。每一步都要通过前边已推出来的另一些关系来证实。通过表示出各种关系之间的联系，同时指明哪些关系能为推出另一些关系做证明而达到了理解。

Perdix的解题过程还表示出了功能理解或有意理解所需要的信息。在每一步，Perdix动作都与系统所表示的一个目标相联系，如图1.1的虚线所示。目标是分等级的，低级的子目标与它隶属其下的高级目标相联系。Perdix就是通过记录下为实现高级目标而提出的每个子目标来达到解题的有意理解的。这种把解题过程表示为一个目标树的方式已成为几种理论系统解题所用的表示知识的一种手段(恩斯特、纽厄尔，1969；威诺格拉德，1972)。

C.知觉和思维过程

几何题是用几何图形呈示出来的。解几何题所需要的许多原理是作为言语命题来学习的。似乎通过对几何问题解决过程的分析可以为视知觉和基于言语原则的思维过程的相互作用提供有用的信息。

对于通过这些分析所获得的知觉和思维过程相互作用的信息

存在着两点争论。一点争论是关于信息形成兼容性这个一般的理论问题。可以认为几何图形给予的信息和以言语命题所表示的信息在某些重要的方面是不一致的，因此，为了使一个问题解决系统能有效地使用二种不同形式的信息，就需要有一个精巧的转换过程。从所进行的分析看来，至今还没有发现在这方面有什么大的困难，这就说明图形信息和言语信息基本上可用同样的形式来表示。第二点争论涉及到问题解决系统在解题过程中利用图形信息的方式。迄今所进行的分析表明利用图形信息一般有三种方式：一种方式是简单的模式识别，用以找出进行某些特定推理所需要的关系；第二种方式是在选择计划时所进行的全局特点的识别；第三种方式是找出图形和通过加辅助线能得以矫正的空间图解之间的差别。

关于表示兼容性的第一点争论，研究者们的主要发现是解几何题所需要的图形信息基本上可以用一个标号网络来表示。许多研究者(帕尔默，1975；温斯顿，1975)已经注意到一个网络可以为表示多种图形信息提供一种便利的方法。回想起来并不奇怪，这种格式为解几何题所用的信息提供了一种完满的表示。而且，把图形信息表示为一个网络也就不需要在系统中进行图形信息和语言信息之间的转换了。应当看到Perdix所执行的模拟是以网络所表示的信息开始的。一个人类问题解决者必须从图形和言语文字中抽出信息来，而这些转换过程在Perdix中没有表示出来。但是，在一个网络表示中两种信息的完全兼容性表明了无论是人类的图形知觉还是言语文字至少都可能构成一种关于成分及其关系的抽象表示。问题解决就是涉及到这些同样形式的抽象关系的控制，而不管信息的来源如何。

第二点争论涉及到问题解决过程中运用图形信息的方式。简单的模式识别，当目前的问题解决计划需要两个角之间的关系或

图形其它成分之间关系的信息时，就要执行模式识别产生式，从而得到了所需要的信息。

第二，图形信息可用于制定计划。具有不同目标的各种计划一般使用了不同的全局情境特点。例如，一个要找出两个角之间定量关系的计划就使用平行边的关系，其它则使用全等三角形的关系。计划系统使用全局特点来选择呈现出先决条件特点的计划。在某些情况下，如果有许多与目标有关的任选成分允许检验的话，那么，计划系统就使用一个扫描过程，在图形中固定一条线并检验图形的有关成分以便找出它们与扫描线之间的距离。

第三，图形信息可用于需要作图的问题。分析表明，图形信息的加工是与详述使用某种计划达到某个目标所需成分和关系的图解表示有关的。这个加工系统能看出图形和图解之间的差别，能通过作图画出与图解相当的图形来。

Ⅸ. 摘要与结论

在这篇文章中，我报告了一个中学几何问题解决的模型。这个模型是一个产生式系统的形式，检验一个数据结构中各种模式的存在，并对这种情境添加上新的有关材料。

这项研究的主要发现是找出了学生解几何问题显然必要的各种知识。所需要的知识有三种：知觉模式识别知识，进行推理的定理知识，提出目标和制定计划的策略知识。

模式识别和定理推理知识是直接使用目前问题解决理论的标准概念来表示的。关于对学生策略知识的模拟提出了一些很有意思的问题并且需要使用模式识别系统来表示一个不确定目标结构、逐级计划系统和作图的图解基模式填充系统。这个模型相当准确地模拟了学生的作业，解释了问题解决中的定势作用。

研究表明对教学的主要意义在于策略知识。推理所用的模式识别和定理推理知识是作为几何课的一部分内容明确地教给学生的。而提出目标和选择计划的策略知识对许多学生来说虽然可以通过隐含着策略原则的例题的教学而得到，但不是作为几何课的授课内容明确地教给学生的。在几何课中增加特定内容问题解决策略的明确教学这种想法有一定的吸引力，虽然这样做可能会有些不利之处。

这项研究还对问题解决一般理论增加些新概念。几何题有些特点被认为是代表了不良结构问题的特性的，不确定目标和几何题作图的分析在一定程度上阐明了良结构问题和不良结构问题解决过程之间的关系。把问题解决表示为模式建构的过程是为把问题解决表示为一种理解过程提供了一种方法。最后，这项研究还表明图形信息和用语言文字叙述的定理的信息是完全不矛盾的，并且提出了解几何题时使用图形信息的一些方面。

这项研究总的结论是现代认知心理学和人工智能的概念和方法显然可用于学校教学任务的分析上。使用这些方法显然已成功而准确地描述了中学几何问题解决作业的一些特点。进一步说，这种问题解决作业的研究也使我们对问题解决一般理论的某些争论有了进一步的了解。

杨 琦 译

二、 何时、何地知晓，如何记忆：一个关于元认知的问题

安 L.布 朗

I. 引　　言

A.这是一个新的记忆发展篇章吗?

有些人初一看到本章的标题可能会想“干吗我们还要再写一章关于记忆的发展呢?”对此我深有同感，并愿由此谈起。近10年来，研究记忆发展的兴趣越来越大。最近在这方面确实有很多文献资料，有的已经撰写成书，有的在一些书里得到了反映，还有不计其数的综述。以后还会更多，对这些雨后春笋般出现的文献资料，若想完成兼收并蓄的任务简直令人望而生畏。由于信息来源太多，要想对这些文献作完整的综述，一定相当长，而且本章也不打算作这种全面的综述。相反，我们在这里讨论的只是比较狭窄的、与元认知有关的那些一般问题解决技能的发展。虽然重点是在元记忆的发展上，但这反映了目前的研究水平，而不是深信元认知技能存在于对人的活动的智慧控制上，同时识记一定不同于无论是通过实验产生的或自然发生的任何存在于其它问题解决情境中的元认知技能。

即使在元记忆的发展这一有限范围内，本章也不打算作完整的评述。因为经过有了一个很好的总结 〔弗雷沃(Flavell)、韦尔

曼（Wellman)，1977)。在本章的回顾(Ⅲ)中代表我们对适当的文献作了最有特色的选择，而这些文献是以我们实验室的研究结果为中心的。主要是着重于学习迟缓儿童在自然的和实验的条件下问题解决技能的效率及其局限性部分。

在这里加以评述的问题解决技能乃是那些在人类和机器智力的许多理论中执行着的技能，如：预报、核对、监控、现实性考验以及对学习或解决问题的各种经过仔细斟酌的尝试协调和控制。我相信在各种学习情境中这些都是进行有效思维的基本特性。有效的思维是智力的一个很好的定义，本章的重点也就是着眼于智力的发展。选择记忆过程来加以论述是因为我们已有了如此广泛详尽的文献。非常简单，人们对正常的记忆技能的发展已经了解得非常详细了，对元记忆觉察的发展的合理描述现在才只是开始。因此，这个信息也可以用来帮助我们理解某些异常的发展。我们对这种传统记忆技能的强调并不等于我们相信存在着一种以某种方式游离于一般智力活动而独立的记忆系统。需要重申雷特曼(1970)的令人信服的观察："记忆行为不仅依赖于记忆的子系统，它反映了作为一个整体的人类认知系统的活动。"实际上，用记忆和记忆术这样的术语只是为了一种方便，这样我们可以和传统的概念一致起来。记忆与其他认知活动是密不可分的这一点今天在现代认知心理学中已确认无疑。而且贯穿本章的那些思维、问题解决、学习等也常常可以由记忆来代替，就是这种认识的反映。因此我们所要集中讨论的是可以在相当广泛的活动中起作用的智力的元认知技能，而不是在传统的记忆这一定义范围内的元认知技能。

B.本书的组织结构

我再次表示同意有些人的意见，他们对本章内容的第二个反

应可能是"如果我们一定还要有这么一章关于记忆发展的话，是否一定要这么长呢？"很显然不一定。可现在它确是这么长。所以我要为不同目的的读者提供一些线索，以说明他们可以选读其中的哪一部分。第Ⅱ部分对"元认知"作了简短的介绍，解释了现在之所以集中研究各种元认知发展的原因。第Ⅲ部分比较长。按照元认知技能的基本结构对我们实验室得到的资料作一评述。这是我们在本章中所要强调的。对我们正在进行的研究方案基本不感兴趣的人这一部分可以略读，并不会影响对后面部分的理解。第Ⅳ部分反映了对最近的一些研究方案中的那些问题类型的关心。这些研究方案主要是强调元记忆术的发展。其中特别强调与教育干涉或矫正教学有关的那些方案。最后，在第Ⅴ部分，考察了经实验室研究和测验证明是合理的许多传统的记忆技能的文化相对性及智力的定义。第Ⅴ部分也讨论了生活条件差的儿童的特殊问题。与此同时还讨论了进一步研究促进迟缓儿童元认知发展的意义。

Ⅱ．元认知是一种副现象吗？

有些人对最近有关认知发展的文献中出现的"元"字感到迷惑不解，并且他们中有些人对元认知表示了认真的关注。所谓元认知实际上是有关问题解决的文献资料中经常研究的材料，只不过最近因为采用了新的提法而显得高贵起来的一种副现象。例如看到元学习、元记忆、元注意、元理解、元语义学等等术语时，表示怀疑的读者可能奇怪为什么非要加"元"字呢？这种附加词是可以解释的。如果能解释的话，它只不过是反映了现实的着重点的转移，仅此而已——我相信就是这么回事。我认为元认知所描述的那些过程是知识的重要方面，主要关心的是人们对自己认知的了解和知

晓而不是认知本身。正象发烧是一种继发性症状、一种疾病的副现象一样，智力评价的结果和对自己认知过程的控制是元认知的基本构成过程的伴随现象。这并不是说所有的认识活动中对自己活动的有意识的控制是必不可少的。我在别处曾集中阐述了偶然学习是一种与有意义的环境相互积极作用的结果(布朗，1975)；在有意识学习以及问题解决这一领域内，对日常活动的有意识的积极的控制调节是智力活动的精髓。我们所看到的那些日常活动所反映出来的威力则是第二位的，是智力活动的伴随现象。

在我们进一步探讨之前，先明确一下什么是元记忆技能是有好处的。同时为了纠正"元"字的滥用，我想引述约翰·弗雷沃(1976b)的一些话，他在这方面所做的工作比其他发展心理学家都多，而且他一直负责这一领域的研究工作。

> "元认知"是人们对自己的认知过程、认知结果及与之有关的一切事物如信息或数据有关属性的学习的认识和了解。例如，有甲乙两种材料，我感觉出乙比甲难学；或者我突然感到需要再核实一遍才能把C作为一个事实接受下来；如果在各类多重选择任务中要决定哪种最好，我们知道要对每种方案都进行详细、谨慎的考察；如果我感到为了不忘记，最好给D作一个注解；那么在这些时候我们就是正在从事着元认知(元记忆、元学习、元注意、元语言等等)。除此而外元认知牵涉到主动监控和连续的调节以及认知对象所基于的那些有关过程之间的和谐结合，通常这些都是服务于某些具体目的的〔P.232〕。

请持怀疑态度的人注意这些活动与传统上作为"学习一技能"所考虑的那些活动的相似性〔布朗和斯迈利(Smiley)，1977a；鲁宾逊(Robinson)，1941〕。元认知的领域并不象它可能表现出的那么

新奇。

科拉尔(1974)提出了是否可以在知识之间作出一定的区分的问题是一个可行的问题：

> 为什么我们不考虑存在着两种形式知识呢？
> (a)关于事物本身的知识，(b)适当地使用它的知识。我对K的了解，其中一个方面肯定是它的本身，例如怎样做乘法。另一方面就是应用这种知识的一系列适当条件。无论哪一方面知识的增加都应该是增加了我对K的理解〔PP.295—296〕。

这一段引文正说明了元认知与其内容方面的基本的相互依存性。这个相互依存性我在后面再讨论(IV.D)。在此我要说明三点：在知识与对知识的理解上适当地使用这个术语表示的这两者之间的差别看来是一种现实的差别，对那些关心什么发展了的人很有启发价值，在教育上，知道什么和知道怎样(布鲁戴Broudy,1977)这两者在传统上的区分仍然是对教育实践(看ⅣF、ⅣG和ⅤD等部分)有着重大意义的一种有效的区分。第二，虽然有关某一过程的元认知及这个过程本身相互之间的界线不清(例如元记忆和记忆)，但是近来被当作元认知技能加以研究的许多技能是超情境的(也就是它们被用于许多形式的问题解决活动中)。最后，如果某人感兴趣的是选来研究的过程的"生态学有效性"，那么元认知技能确实在"现实世界日常生活"的情境中有着可认出的副本。依据某些有效性、经济性及常识的真实性等标准来核对一个操作的结果就是一种元认知技能。解数学题，识记一段散文，按菜谱烧菜或者装配一辆汽车，无论什么任务都用得上这种元认知技能。在解决问题时，对自己最近的知识状态进行一下自我讯问，这无论是在实验室、学校或日常生活中都是一种最基本的技能。

总之，为深入地研究而把元认知技能人为地分离出来是可行

的，它将帮助我们集中注意于传统的认知领域中的类似性而不是差异性(弗雷沃，1976a)。此外，由于元认知要求有内省自己的作业并把自己的观点与别人的观点相区别的能力，因而有关的研究领域，例如社会认知，角色扮演和交流也就变得直接有关了。另外，因为对自己的作业的自我评价不可能是客观的——例如自我讯问常常要受到自己自我胜任能力感的影响——这样以前被分离出来的有关个性发展的那些领域显然也互相联系起来了(如，害怕失败、对成就的需要、外部与内部的控制、掌握的无望、抱负水平)。由于强调了元认知的发展，不仅在传统的认知领域内的人为的割裂会被削弱，而且发展心理学所研究的各个独特领域之间的那些界线也可能成为问题。如果我们认真地把发展着的儿童看作是一个完整的人，而不是处于各个发展阶段上的某些有趣的零碎技能的贮藏所的话，那么这些界线的重新排列是不可避免的，而且是有利的。

Ⅲ. 元记忆：关于文献的选择性综述

大多数关于元认知的实验研究集中在元记忆上——关于自己记忆能力及策略的知识。这个术语是弗雷沃(1970)引入的。他和他的学生一起对于元记忆知识的发展积累了丰富的资料。对这些材料的卓越综述早就有了，(弗雷沃和韦尔曼，1977)我们在此不再赘述。所以，在这部分要介绍的是我们实验室中新近的一些研究，主要是关于那些学习比较迟缓的儿童的元记忆知识。

此外，这一节还要集中讨论目前的发展研究中一个带有普遍性的问题：为了进行研究所选择的特定形式的元记忆使得对其中所包含的操作活动的复杂性估计不足。在这些研究中主要的注意力一直是在元记忆知识的孤立的、片断的内省上，而不是关于个人、任务和策略等这些变量(弗雷沃和韦尔曼，1977)之间的复杂的

连续的相互作用上。而正是这些变量才在实际的审慎的记忆的尝试中起作用。所需的内省的难度水平这个问题一直没有得到充分的考察。因此，如果很快地细读一下现存的发展研究的文献可能会认为到三年级时元记忆知识就非常成熟了(看Ⅳ、C部分)。但是我要说明这正是由于至今所考察的还都是些简单的元记忆知识造成的一种错觉。

如果考虑一下在许多流行的记忆模型中称为监督者的关于中央处理机、解释程序和执行程序的有关操作，那么我们就可以对成熟的记忆者所需的元认知能力的复杂性提出一些想法。能够对本身的操作作出智力评价的能力是当前许多理论所热衷的中枢机制的一个基本特性；对本身工作的某种形式的自我意识或明确的了解是有效的问题解决系统的一个关键〔贝克尔(Becker)，1975；鲍布罗(Bobrow)，1975；鲍布罗和诺曼(Norman)，1975〕，对执行程序的这种基本要求可以说明问题的复杂性。它必须包括这样的能力：1)预报系统的容量限度；2)知晓系统的探试程序的指令系统和这些探试程序的适当应用范围；3)识别出身边的问题并描述其特点；4)设计和制定适当的问题解决策略；5)监察和监督系统所调用的那些程序的效果；6)在面临成功或失败时，能动地评价这些操作以便能机智地安排策略活动的终止时间。做出这些形式的执行决策也许是有效的问题解决的关键。因为在适当的时间和适当的地方使用适当的知识或者程序去获得知识是智力的精髓。

显然，一个儿童是对他自己的记忆有所了解还是对问题解决过程有所了解，这有赖于所进行的判断的复杂程度。例如，如果考察更复杂的协调和预报，那么在迄今为止的元记忆研究中，三年级似乎就表现了成熟理解的情况就不复存在了(布朗，1977a；布朗和德洛切正在排印中；巴特菲尔德Butterfield和贝尔蒙特Belmont，1977)。在下面的综述中，我们打算根据所进行的判断的复

杂程度来考虑最近的一些文献。

A.副无知：不知道你何时知道或知道什么

在所有的记忆和问题解决任务中所包含的一个非常基本的自我觉察形式是认识到有一个你知道什么和你不知道什么的问题(布朗，1975)。我们非常感激希伯(Joan Sieber 1968)促使我们注意到"副无知"这一问题，即使我们要花一点时间才能懂得这一观察的重要。希伯是在这种情况下使用这个术语的，当一个人意识到自己无知时，那么他就不是无知的了。霍尔特(Holt)在其《儿童怎样失败》(1964)一书中所写的敏锐的观察对这种情况作了很好的说明：

> 一个好学生应具的一种品质是学会了解自己的心理和自己的理解程度。他可能常常会说他不理解，只是因为他在不断地检查自己的理解。差等生可以说就不去注意他是如何理解的。他们在大多数时间都不知道他自己是理解还是不理解。所以问题不是要学生来问我们他们不知道什么，而是要使他们意识到他们知道什么和他们不知道什么之间的区别〔PP.28—29〕。

1.元理解。 弄清楚自己是处于无知或启蒙状态，这就是一种元理解。理解各种指令就是一种理解信息的例子，而知道他是不是已经理解了就是一个关于元理解的例子。霍尔特(1964)关于儿童对许多学校问题迷惑不解的那些清晰的描述包含了许多元理解失败的例子。例如，叫一个孩子把以P结尾的动词排列起来，这时她心烦意乱地重复"我不会"，可她总也说不出为什么不能理解。然后，霍尔特问这个孩子知道不知道什么是动词；还给了她一些例子。困难排除了，孩子继续做下去。霍尔特(1964)相信这个孩子

不会问什么是动词，就是因为：

她不了解自己不知道什么。她所知道的就是叫她开始做某一件事，而她并不知道要做什么。她不能分析这些指令，不能找出其中哪一部分有意义，哪一部分没有意义，在这里她陷入了迷茫，变得无知〔P.145〕。

虽然霍尔特在他那本令人陶醉的书中写下了有关儿童难于评价自己知识状态的许多轶事，但是关于这种发展上相关现象的控制实验测验却还不多。在最近的一系列研究中，马克曼（Markman 1977)考察了低年级儿童对自己不能理解的不敏感性。让一年级到三年级的孩子帮助实验者为一种新游戏(纸牌游戏)设计一套指导语，这个指导语要用来教给和他们相同年龄的其他孩子。指导语显然是不完整的。用孩子是否需要更多的信息来测量其是否认识到了他不能理解。例如，实验者和儿童每人拿到四张字母卡片，给儿童这样的指示："我们每人都把牌放成一堆，翻开自己那堆顶上的第一张牌，看谁有那一张牌，然后我们再翻这堆里的下一张牌，这次再看谁有那一张牌。最后谁的牌多，谁就赢。"没有说哪一张牌是那一张牌。这些儿童不仅需要许多额外的提示才能说他们不理解。而且也看到他们需要重复提示，甚至于他们还不知道该怎么说，他们就想要执行这种任务了。因为监控自己对无论是书面还是口头的指令和信息的理解能力是所有问题解决技能的基本的先决条件。所以应该欢迎在这方面的深入研究。

2.知识推理的缺乏。 虽然，好象是觉察到你知道它是什么或不知道它是什么是一个更复杂的元理解形式的早期预兆，但这种内省的困难是一个重要问题。因为在一定的条件下，即使大学生在估价其本身的知识状态时也可能有问题。查明自己不知道什么或可能不知道什么的过程也许包含了十分复杂的推理形式，成

熟的问题解决者不仅对所知道的事实的可理解性有合理评价，而且他们也能认识到哪些事实是不可能知道的，以及什么东西是可以在他们已知的基础上推断出来的。成年人立即就能认识到不可能知道查理斯·狄更斯的电话号码(诺曼，1973)，而且他们是通过推论得出这个结论的，这样的推论自然要涉及到他们已有知识的其它方面。儿童不仅比成人知识少，而且他们的知识常常组织得不好，不完整，不一致，他们也缺乏成年人那种能从不完整的和矛盾的知识库中推出信息的复杂的推理系统。科林斯(Collins，1977)已经表明大学生使用各种推理策略来评价他们已有的信息在多大程度上可以说明一个假设的可能性。在此不详述科林斯的工作，而且所描述的这种推论有很多种类型。不过我希望读者应把他的苏格拉底式的对话记录看作是有关这类技能复杂性的非常丰富的实例。

3.专家。 元理解问题包括觉察出你不理解和觉察出策略监控，以便断定你有完成某一任务的最佳信息。考虑一下关于一个有意记忆任务的问题。记忆者必须再认出这一类问题，并且认识到这些问题需要他们的适当活动。仅就识别出需要有意记忆的那种问题的种类就可能给幼儿的元认知能力造成明显的负担。那么熟练的记忆者是怎样来识别特殊记忆任务中所涉及到的问题呢？

> 没有自我顾虑的记忆专家可能不计较心理学家进行大多数记忆实验的那种方法。专家可能要问“我必须记什么？”“多少项？”“有多少时间？”“测验的性质是什么？”等等。他们要知道他们需要知道的一切以便理想地完成任务——而且一定要到什么都不缺了才满足。〔布兰斯福特(Bransford)尼奇(Nitsch)和弗兰克斯，1977，P.38〕。

比奈(Binet)用“聪明的计算者”和“杰出的记忆术者”与棋手

所作的开拓性研究〔比奈，1894；里夫斯(Reeves)，1965〕也说明了专家在着手进行工作前不仅需要充分地了解一个问题的各个方面，而且也喜欢用一种最佳方式把输入信息组织起来以求实用。辨别问题的确切的性质可能是一个复杂的活动。

因此，记忆任务可能包含了能使大学生的智力都感到是负担的复杂的元认知内省。不过还是让我们看看那些幼稚的专家——前运算阶段的儿童吧。我们有什么证据可以说儿童对于甚至是"简单的"记忆任务知道些什么呢？获得这种信息的一种方法就是问他们是怎么确定答案是正确的(也就是获得了自信心的等级评定)。

4.自信心。　许多记忆范例的一个特点是可以断定被试对他们反应的正确性充满了信心。所以，在连续的再认范例中，人们可考虑把命中率(正确再认的项目)与假报率(错误再认的项目)相比。保守的反应倾向就是属于这种情况：如果被试指出一个项目是以前见过的，而且它也确实是以前见过的，那么被试的假报比率非常低。儿童有一种保守的反应倾向〔布朗和坎皮尼(Campione)，1972；布朗和斯科特，1971〕，表现出他们能"老练"地识别出以前曾经看见过的项目。这是关于信心的一个非常粗糙的测量，d′测量已被特别引入来区分出再认作业的保持基成分，使用一个d′测量把我们的再认作业中的保持基成分与标准基或决定基成分区分开。珀尔马特(Perlmutter)和迈尔斯(Myers)(1974)以及伯尔茨(Berch)(1975)使用d′测量也得出结论说非常幼小的儿童在再认作业中采取一种保守的反应倾向。另外，也有可能直接解决问题，并且要求被试去评价他对一个特定量表所做的"老的"或"新的"反应的信心。伯尔茨和埃文斯(1973)在6岁儿童中成功地应用了这个程序。虽然从判断自己再认作业的准确性方面来说，9岁儿童是比6岁儿童好，但是在某种程度上，较小的儿童也有能力评价自己反应的准确性。在判断一个老项目是否属于"老项目"时，儿童

的信心水平越低，这个项目属于“老项目”的可能性就越低。

下一部分我们将充分考虑儿童相识感的经验，不过在布朗和劳顿(Louton 1977)的一个研究中提出来的关于信心评价的一个因素，是我们在这里所感兴趣的。可教育的智力迟钝儿童，不管在一个再认选择之前是否能预报出他们的相识感，但是在做出再认选择之后他们能够评价自己的反应准确性。表2.1给出了关于“确实”和“不确实”判断的再认准确性资料。在年幼组内各类间的差别比较低，即使这样它也是可信的。

莫伊纳汉(Moynahan)(1976)也报告了作业前预报的可能准确性和作业后对准确性的自信心之间的差异。给一年级和三年级孩子学习八项表。在回忆试验之后紧接着要求他们指出已经回忆了多少个项目。在我们实验室中用一些可教育的智力迟钝儿童做了一个类似的研究。在估计他们已经回忆起了多少项目时，所有被试几乎都很少有错。虽然这个年龄的儿童在尝试回忆以前还不敢预言他们的估计（布朗，坎皮尼，和墨菲Murphy，1977）。另外，在我们的研究中还要求他们指出回忆出的是哪一个项目。儿童又一次真的在这个任务中错误比较少，他们回忆的项目有95％可以被正确地确定出。显然，甚至年幼的孩子也可以准确地评价他们先前的再认准确性和回忆作业的成功。如果情境是一个比较直接的再认记忆任务，他们能意识到他们什么时候知道和知道什么东西。

表2.1　在相识感的实验中准确性与关于再认选择判断自信的关系

被　试	小(心理年龄6)	中(心理年龄8)	大(心理年龄10)
N	17	15	27
确　实	0.54	0.71	0.66
不确实	0.34	0.34	0.28
差　异	0.20	0.37	0.38

注：引自"可教育的智力迟钝儿童的相识感的实验研究"A.L.Brown和S.L.Lawton, Developmental Psychology,1977,13,364—370,1977年美国心理学会翻印。

B.预报

反应后能准确地评价作业与试图检索前能预报准确性形成了鲜明的对照。莫伊纳汉(1976)、布朗及劳顿(1977)认为预报比较难是由于这种活动所包含的抽象程度所致。在反应之前预报需要对那些尚未发生的认知活动有想象能力。有很明显的证据表明这样的活动对于幼儿来说要更难一些。

1.相识感。首先考虑相识感这种经验。和缺乏知识的推断情况一样〔科林斯，沃诺克(Varnock),艾劳(Aiello)和米勒(Miller),1975〕，在知识中存在着的缝隙和用来填补"那个强烈活动着的缝隙"的主动的策略尝试之间做出区分是必要的。这类主动检索的尝试一直是在欲言难吐的实验中加以考虑的情境。在成人中，欲言难吐的现象首先是由相识感组成的，这是一种被称作后找信息并完全是等待着适当接近的感觉，紧接着便是进行积极的策略尝试以促进对那些虽然一时不可能接近但被推定最终是有用的那些材料的重新检索。

布朗和劳顿(1977)对处于各种认知能力水平的心理年龄6—10岁可教育的迟钝儿童进行了两个关于相识感经验的研究。第一

个研究类似于成人所用的范例。事先训练儿童用1—3个他们能再认出他们不能回忆起的熟悉图形当筹码做为赌注来打赌。因为大部分年龄大的儿童回忆得比较准确，所以只能用年龄比较小的被试来考察他们的相识感。除了由于准确性的问题大批被试退出之外，我们怀疑其余相当数量儿童的数据，因为他们好象使用了不适当的打赌程序。这些儿童几乎每次试验都赌三个筹码而且总是正确的。以后的询问发现他们存心说不能回忆熟悉图形以便他们能享受打赌程序的乐趣。此外，其中有几个人虽然没有意识到他们的回忆是否正确，但是他们确信如果正确回忆不多给的话至少总能得到三个筹码。他们认为他们赌三个赢三个，他们差不多获得六个。在下面的分析中不包括这些可疑的数据，在67个被试中最后只剩下13个心理年龄6岁和14个心理年龄8岁的儿童。

如果只考虑"可接受的"数据，那么被试赌一个、两个、三个筹码所得的正确回忆的条件概率见图2.1。较大的儿童赌得越多的时候他们确实回忆得越好。较小的孩子赌一个或两个筹码没有什么差别，但他们赌三个筹码时其正确再认的概率是高的，这表明在较小的孩子中某种相识感的敏感性。

从打赌程序中体会到的问题和因年龄较大的儿童的很高的回忆成功率所造成的相当大的退出率以及这些结果的解释都必须小心地加以重新处理。由于这些原因，第一个实验被重复，用了不象赌博程序那么有吸引力的一些东西作为预报的测量。向心理年龄6岁(N=17)，心理年龄8岁(N=15)和心理年龄10岁(N=27)的三组可教育的迟钝儿童，显示100张名人照片，要他们说出正确的名字。记下每一张儿童不能说出名字的照片并用于他的相识感测验中。在每次相识感试验中，让被试看到一个他先前不能命名的人物并要求他或她指出他们是否能够再认他的名字，如果能再认出他的名字，就说能(Yes)；如果不能再认，就说不能(No)；如果可能

再认，就说或许(Maybe)。然后，读四个名字，一个是对的，一个是完全不可能存在的名字还有两个是从和目标相同的类别中取来的名字。例如，如果理查德·尼克松(Richard Nixon)是目标，那么三个使人迷惑者将是约翰·肯尼迪(John Kennedy)，阿伯拉罕·林考姆(Abraham Lincoln)和罗纳德·麦克唐纳 (Ronald McDonald)。

被试预报他能、不能、可能再认名字时作出的正确反应的条件概率和第一个实验的数据一起呈现在图2.1中。年龄最小的一组没有预报出他们以后能再认出名字。不过，心理年龄8岁和心理年龄10岁的两组在他们预报能够再认时，真的都确实识别出了更多的项目。

对这两个实验的数据所作的分析表明了在不能回忆时关于预报再认准确性能力的发展趋势。对最年幼的儿童的敏感性的估计上有一点混乱。在第一个实验中，他们确实表现出具有预报再认准确性的能力，因为他们在三个项目的打赌试验中确实比较好，可是在第二个实验中，比较有代表性的年幼儿童的样本，在他们预报能够再认出名字时，并未能再认出更多的项目。这两个实验之间的差别表明了只用一种任务来评价儿童的元记忆能力时可能存在的问题。考虑到他们在各种任务，各种情境中的作业才能对他们的能力有一个更为准确的了解。

2.广度估计。 已经受到明显注意的另一种形式的预报是年幼儿童估计记忆能力的估计能力。弗雷沃，弗里德里克斯(Friedrichs)和霍伊特(Hoyt)(1970)让二年级、四年级的儿童和幼儿园的儿童一起来估计一下他们的回忆广度。一张、二张、三张一直到十张这样递增地向他们呈现一些图片，儿童的任务是看过不同张数的图片系列后，要说出是否能够回忆出张数最多的那个图片系列。半数以上的托儿所和幼儿园儿童预报能够回忆起全部十张图片，相反大一点的孩子只有少数预报他们能够回忆全部十张图片。如

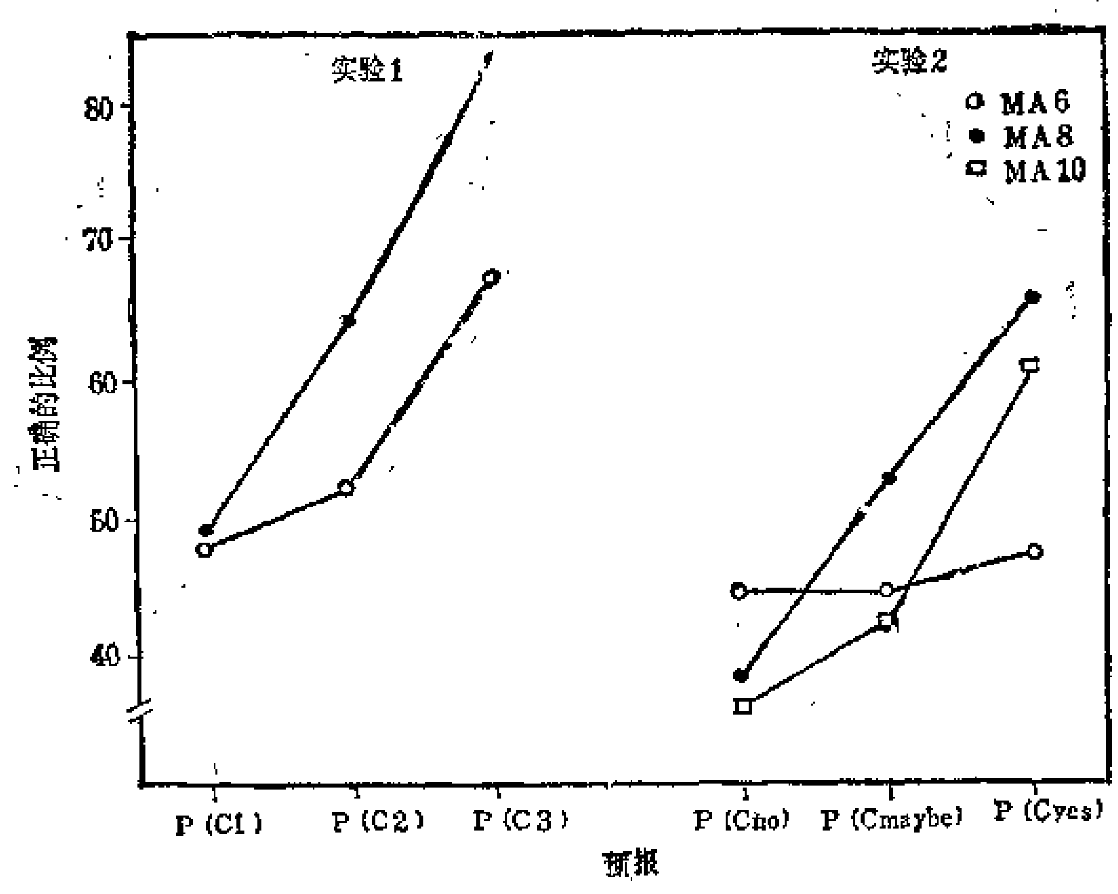


图2.1　实验1中赌1、2、3个筹码的正确认知概率和实验2中作出能、可能、不能的正确认知概率。

果我们把实际广度±2这样一个估计值作为一个现实估计值的测量，这个组平均数符合二四年级水平的标准，但对年幼的儿童不是这样。

马克曼(Markman)(1975)和约森(Yussen)和利维(Levy)(1975)象布朗等人用可教育的迟钝儿童所做的实验那样重复了这些结果，所有三个研究都涉及到一个训练因素，以后我们还要讨论(Ⅳ，F部分)。这里我们所感兴趣的是在布朗等人的研究中所碰到的问题：在被试指出所给的图片张数超出他们的广度时，实验不停止，相反实验者继续一项一项地显示那些图片，一直到所有十张图片都显现出来为止。程序上的这个小小的变化结果产生了三组被试的不同反应：一组被试判断是现实的(实际广度±2)，另一

组被试象先前的研究一样判断是不现实的，还有第三组被称作是反复无常的(32%)。第三组儿童表现出某种张数的卡片对他们是太困难了，可是后来他们又能估计一个张数较多的系列。允许被试可以有这种反复无常的反应，明显地改变了所得结果的情况。目前还无法了解，如果给以机会的话，先前的研究中有多少"现实的"被试会产生反复无常的反应情况。例如，如果我们在被试作出第一个十分困难的反应后就停止的话，那么我们研究中的80%的心理年龄 6 岁的反复无常的被试判断为是现实的，这种程序以前用过。就总体而言，这将导致这样一个估计值，即有28%的心理年龄6岁的被试被判断为是现实的，这个估计值和马克曼和弗雷沃的估计值不同(对幼儿园儿童来说是22%和36%)。这个结果说明把幼儿的口头反应当作其元记忆能力的真实估计时所产生的问题，还有一个困难，后面我们会谈到它(Ⅳ，E部分)。

3.估计任务难度。 在关于幼儿对任务难度的觉察方面有一些资料是很有用的。我们首先考虑克罗泽尔(Kreutzer)，伦纳德(Leonard)和弗雷沃(1975)所提供的问卷资料。克罗泽尔等人向儿童提许多问题，这些问题涉及到他们对自己及别人的记忆过程的了解程度。我们对可教育的智力迟钝儿童用同样问题(稍加修改)作为对元记忆任务(有经验的被试和首次参加实验的被试)一致性训练二年的一般效度的测量。我们将在后面再来谈训练的问题(Ⅲ，E)，但是在适当的时候，我们还要谈到迟钝儿童的那些资料。我们选择克罗泽尔等人的三个项目作为儿童任务难度觉察的特别好的测验：编排故事问题，任意对立项目和机械释义问题。

设计编排故事的问题是为了考验儿童是否具有这样的知识，即把一些要记忆的项目安排成一种叙述关系以便从中收到效果。结果见表2.2。正常儿童的反应类型中存在着明显的发展差异。只有50%的幼儿园儿童体会到这种编排比较容易。而所有三年级儿童

都认识到了这一事实。另外，70%年龄较大的儿童对于他们所作的选择可以说出正当的理由，但年幼儿童大约只有15%能够这样做。

任意对立项目是预报任务难度的另一个例子。这是要求儿童去判断学习一系列随机配对的字词或根据共同联系(对立的)配对的字词表的相对难度，这些结果见表2.3。较年幼的儿童准确地预报有困难，但是到三年级就可以做出比较准确的预报了。心理年龄8岁的儿童只有一半能够准确地预报。如果同时考虑儿童为自己的选择所说明的理由的话，那么就更进一步说明了较年幼的儿童有困难。因为正常的幼儿园儿童和心理年龄6岁的智力迟钝儿童只有少数能说出正当的理由。

表2.2 编排故事[a]：做出各种选择并作出适当说明的被试的比例[b]

被 试	故事形式				提出适当的理由
	N	容易	难	其它	
心理年龄6岁(MA 6)					
正常幼儿园[c]儿童	20	0.50	0.25	0.25	0.15
首次参加实验的迟钝儿童	21	0.81	0.19	0.00	0.05
有经验的迟钝儿童	40	0.78	0.18	0.03	0.20
心理年龄8岁(MA 8)					
正常三年级[c]	20	1.00	0.00	0.00	0.70
首次参加实验的迟钝儿童	28	0.75	0.04	0.21	0.04
有经验的迟钝儿童	30	0.80	0.20	0.00	0.40

a.取自克罗泽尔等人的问卷(1975)

b.评定者间的信度=0.96

c.正常儿童的资料取自克罗泽尔等人的材料(1975)

表2.3　　任意对立项目[a]：做出各种选择的被试比例[b]

被　试	N	对立的容易的	任意的容易的	相同的	提出适当的理由
心理年龄6岁(MA6)					
正常幼儿园[c]儿童	20	0.30	0.50	0.20	0.10
首次参加实验的迟钝儿童	21	0.43	0.57	0.00	0.10
有经验的迟钝儿童	40	0.31	0.69	0.00	0.12
心理年龄8岁					
正常三年级儿童[c]	20	0.90	0.10	0.00	0.65
首次参加实验的迟钝儿童	28	0.57	0.39	0.04	0.42
有经验的迟钝儿童	30	0.48	0.52	0.00	0.40

a.取自克罗泽尔等人的问卷(1975)

b.评定者间的信度＝100%

c.正常儿童的资料取自克罗泽尔等人的材料(1975)

所选择的第三个问题是机械释义问题。克罗泽尔等人的记分和制表是极为复杂的，我们仅选择某些突出的特点来了解。向儿童问一些关于学习一个录音故事的难易的问题，看他们是否理解差异学习、要点回忆的要求与逐字回忆的要求。主要结果见表2.4。我们看到一个典型儿童问实验者，是用她自己的话，还是用故事里的话来回忆。然后向所有的实验被试提出下列问题："你认为她为什么要问这个问题?"反应数字表明逐字回忆的觉察是更困难的，这在表2.4的第一列给出了。第二个问题："这个问题会帮助她得出回答吗?"做出肯定回答的反应见第二列。问题3和4基本上是要求儿童指出如果他们要逐字地学习或者要得出要点的话，他

们该怎么做。问题5和问题6是问儿童"要点回忆"和"逐字地回忆，哪一个比较容易。用自己的话来回忆是比较容易的，见第五列。有充分正当理由的见第六列。这种预报与前面的两个项目相比可以看出同样的倾向。大多数儿童指出用自己的话回忆是比较容易的。不过大多数年龄较小的正常儿童和可教育的迟钝儿童不能为他们的反应说出正当的理由，也不能指出符合每种回忆要求的适当的学习活动。

这些问卷资料表明随着儿童的成熟，他们就越来越能觉察出任务的难度。正常8岁儿童对我们这里所考察的任务难度问题是非常敏感的，不过可教育的迟钝儿童的进步却并不是这样惹人注目的。有趣的是在这种游戏中，两组儿童都容易产生错误印象。60%的6岁正常儿童和55%的8岁正常儿童相信彩色图片比黑白图片更容易学。可教育儿童也是如此，70%心理年龄6岁的儿童和55%心理年龄8岁的儿童也有类似的错误印象。

儿童判断任务难度的效率客观地依赖于他们所必须判断的任务类型。在问卷的研究中，几乎所有智力迟钝儿童都体会到再认是比回忆更容易的一种任务，同时几乎没有一人对倒摄干扰有所评价。这一点再次表明了所欲作出的判断的难度和觉察的相互关系。年龄较小的儿童不单是觉察出或觉察不出任务难度。对有些类型的问题他们很快地就觉察到了，相反对另一些类型他们则是完全不能理解的。

任务判断复杂性的这种影响在实验研究中以及问卷资料中可以看到。首先，我们来考虑一个估计逐字学习的项目表难度的研究。

坦尼(Tenney1975)要求幼儿园的儿童，三年级和六年级的儿童用易于回忆的方法把一些单词排列成表。把他们设计的项目表和自由联想项目表作比较，以此来评价组织的策略。年龄较小的

表2.4　　机械释义问题[a]：做出各种选择的被试的比例[b]

问题号码		1	2	3和4	5	6
被　试	N	适当的反　应	肯定的反　应	适当的活　动	自己的话容易	提出适当的理由
心理年龄6岁（MA 6）						
正常幼儿园[c]儿童	20	0.05	0.50	0.20	0.55	0.10
首次参加实验的迟钝儿童	21	0.00	0.90	0.00	0.76	0.00
有经验的迟钝儿童	40	0.05	0.95	0.05	0.82	0.08
心理年龄8岁（MA 8）						
正常三年级[c]儿童	20	0.55	0.75	0.70	0.90	0.75
首次参加实验的迟钝儿童	28	0.21	0.82	0.18	0.64	0.28
有经验的迟钝儿童	30	0.20	0.83	0.30	0.73	0.20

a.取自克罗泽尔等人的问卷(1975)

b.评定者间的信度＝0.96

c.正常儿童的资料取自克罗泽尔等人的材料(1975)

儿童在两种条件下基本上编排了同样的项目表，年龄较大的儿童则倾向于按类组织。

作为我们对可教育儿童元记忆情况纵向研究的一部分，我们用心理年龄6岁和心理年龄8岁的儿童与正常班级的儿童重复了坦尼的研究，在实足年龄和心理年龄上与可教育的样本相匹配。象坦尼的研究一样，儿童选出了一些含有4个单词的单元：关键词→及他们认为借用关键词就比较容易回忆的另外三个词。各类单元结构的详细情况见表2.5。经常发生的三种组织是：分类学的

类别、主题类别(也就是围绕着一个主题把一些对象组织在一起：婴儿、儿童小床、拨浪鼓、尿布)以及押韵和同声类别。其他所有的反应都是随机出现的。在正常的总体中，随机反应的发生率象声音反应一样随年龄的增长而降低。而按类别组织的反应则有相应的增加。可教儿童中没有发现随心理年龄的增长而有所改善，并且还产生许多随机的单元。有趣的是主题反应的发生率非常高。虽然丹尼(Denney)和杰奥布鲁斯基(Ziobrowsk1972) 发现年龄较小的儿童更多的是按主题而不是按分类学来组成单元的。坦尼(1975)在她实验的任何一组被试中都没有发现显著的主题反应。如果我们考察所有分类反应的比例（除随机反应之外)，在所有各组中主题反应都有很高的比例(正常被试：实足年龄 6 岁＝0.51，实足年龄8岁＝0.26，实足年龄10岁＝0.31；可教被试：心理年龄 6 岁＝0.36，心理年龄 8 岁＝0.30)。正常幼儿园儿童中主题反应占了所有各类反应中的一半；其他各组被试使用这类反应在所有编造的单元中大约占1/3。

表2.5　　产生回忆表的学生的分类

被　试	N	类　　别			
		分类学	主题	押韵/同声	随　机
正常组					
实足年龄6岁	20	0.21	0.37	0.14	0.28
实足年龄 8 岁	20	0.72	0.26	0.01	0.01
实足年龄10岁	20	0.68	0.30	0.00	0.02
可教组					
心理年龄 6 岁	27	0.38	0.27	0.10	0.29
心理年龄 8 岁	31	0.42	0.23	0.12	0.23

坦尼用来表明儿童对任务理解的另一种测量是儿童所形成的

类别是宽广的还是狭窄的。所谓狭窄的类别是指在单元内部的组织形式本身可以成为一个有助回忆的附加线索。例如，星期一、六、四、二，可能是分类上比较宽广的反应，而星期一、二、三、四则是比较窄的一种反应。显然，狭窄的反应象(1、2、3、4)是一个容易列表的较好的例子，所以我们要看类别反应中宽广和狭窄反应的比例。这些数据见表2.6。这里又一次表现出了发展的倾向。狭窄的分类其数目随正常儿童实足年龄的增长而增加。可教儿童也表现出有所改善，但他们的作业比起其它正常组来更象正常的幼儿园儿童。如果我们只看狭窄分类反应，最策略的反应形式发生的比例，那么这个发展模式就看得更清楚了。实足年龄为6、8、10岁的正常儿童这个比例分别是0.27，0.75，0.80，心理年龄为6岁和8岁的可教儿童分别是0.21和0.33。大多数正常的10岁儿童能够"理解"并用狭窄的分类作词表，与此同时在发展不太成熟的孩子中这类有效的反应是不太多的。

表2.6　　在类别反应中宽广和狭窄反应的比例

被　试	类	别	
	N	宽	窄
正常组			
实足年龄6岁	20	0.62	0.38
实足年龄8岁	20	0.25	0.75
实足年龄10岁	20	0.18	0.81
可教组			
心理年龄6岁	27	0.67	0.33
心理年龄8岁	31	0.49	0.51

儿童对任务难度估计的另一些研究是莫伊纳汉(1973)和萨拉塔斯(Salatas)及弗雷沃(1976)的研究。这些研究是要儿童回答是

分类的词表还是不分类的词表比较容易回忆。关于这些研究我们要在预报和作业之间的关系那部分再来考虑(Ⅳ.D)。在此我们仅指出发现了类似的发展倾向就够了：年龄小的儿童比年龄大的更困难一些，但是到三年级就几乎都可以做出准确预报了。可教儿童的结果与正常的幼儿园儿童类似。

我们想在这一部分描述一个深入的研究，因为这个研究表明如果用较复杂的材料，那么对任务预报的这种发展倾向可能表现在更晚的年龄阶段中。研究的程序是以约翰逊(Johnson)的研究为基础的，他设计了一种把一段散文划分成一些有意义的小单元的客观方法。一段散文经这样划分以后，分别由一些评定者来判断这些单元对表现故事主题的重要性。结果发现更加无关的大学生样本的回忆分数是由所评定的单元的重要性决定的。约翰逊的程序尤其适用于儿童，因为它既提供了一种从一段散文中能回忆出什么东西的定性的方法也提供了对元理解的测量。这样不仅可以查明儿童的回忆模式是否与成人相类似，而且也可以看出儿童是否对课文材料有足够的了解，从而能够决定哪些是重要的单元。

让8—18岁之间的学生去评价或回忆民间故事。所有要评定或回忆的故事，事先都由若干组大学生对其重要的单元做过评定。对12和18岁学生所用的程序类似于约翰逊对成年人所用的程序。这些被试分成组，先听一段故事录音，同时还从头到尾读一遍。这个故事按照先前确定的段落单元，每一单元印一行。在第二次阅读后，告诉被试各个单元对整个故事的重要性是不同的。如果不破坏故事的主题，有些不太重要的单元可以省略。首先指令他们省略大约$\frac{1}{4}$的单元，即他们认为最不重要的，用蓝铅笔把它们勾掉。然后要他们用绿铅笔删去$\frac{1}{4}$的单元，用红铅笔再删掉$\frac{1}{4}$的单元。最后只剩下最初所给的全部单元的$\frac{1}{4}$。这个程序产生了重要性判断的四种水平。开始删去的项目(最不重要)其重要性的分

类为1，留到最后的(最重要的)那些分数为4。年幼被试的程序基本相同，只是他们是一个一个看的，并且对评定程序明显地作了预先的训练。

四个实验组的一般重要性的评定等级与大学生以前评定的等级相比较。这些数据见表2.7。

表2.7　四个实验组的一般重要性评定等级与预先判断的重要性等级的比较

被　试	重要性水平			
	1(最不重要)	2	3	4(最重要)
8岁	2.41	2.52	2.51	2.56
10岁	2.42	2.35	2.46	2.76
12岁	2.02	2.36	2.58	3.05
18岁	1.61	2.09	2.78	3.52
总　计	2.12	2.33	2.58	2.97

注：引自"散文的结构单位的重要性评定：元认知发展的问题" A.L.布朗和斯迈利 Child Development, 1977, 48, 1—12，儿童发展研究会1977年翻印。

在所研究的整个年龄范围内，可以看到有一个明显的发展趋势，那就是在所研究的所有年龄上都表现出对所评定的重要性程度的敏感性不断提高。只有18岁组才能可靠地区别出四种不同水平的重要性。12岁组不能区分出两种中间的重要性水平。18岁组关于四种重要性水平的平均分数的值域比较大，这也说明对重要性精细程度的敏感性在中学阶段应该继续提高。8岁组在他们的评定中区别不出各种不同的重要性水平。即使是10岁的学生也只能从各种水平中区别出最重要的。所以在考虑课文段落的构成单元的重要

性时，大学生甚至于12岁各组之间存在着明显的一致性，但8—10岁的孩子不可能根据在课文中的重要程度区分出更小的单元。

所有的儿童在回忆各段落时，也在评价这些段落。各年龄的回忆分数对成人评价的各单元的重要性水平是极为敏感的。在表2.8中给出了这些数据。虽然年纪大的比年纪小的回忆得要多一些，但是在8—18这个年龄范围内一般结果是一致的。最次要的单元比其它单元回忆得更少。最重要的单元回忆得最多。

表2.8 作为年龄和结构重要性的函数的正确回忆的平均比例

被　试	重　要　性　水　平				
	1(最不重要)	2	3	4(最重要)	总计
8岁	0.17	0.22	0.38	0.61	0.35
10岁	0.23	0.32	0.48	0.68	0.43
12岁	0.28	0.39	0.51	0.75	0.47
18岁	0.27	0.39	0.54	0.74	0.48
总　计	0.23	0.33	0.48	0.69	

注：引自"散文的结构单位的重要性评定：元认知发展的问题"A.L.布朗和Child Development, 1977,48, 1—12,儿童发展研究会1977年翻印。

我们还有另外一些补充资料表明结构重要性的这种影响也为比较年幼的儿童所发现。实验者给37个托儿所的孩子和20个幼儿园的儿童也讲了上述同样的故事，要求他们在各种条件下去回忆，这些条件我们就不必在这里考虑了。虽然要把孩子们回忆的谈话记录下来是很困难的，而且他们能回忆起来的单元的数量是很少的。不过我们的确发现他们对重要程度的敏感性模式是和大学生评定的结果一样的。在32个谈话记录中，至少能回忆出15%。按四种重要程度来说其回忆的平均单元数分别为0.12,0.08,0.22和

0.42。水平2、3和水平3、4之间的差异是可靠的。这里再次表明最次要的单元(水平1和2)即使能回忆的话也是很少的。同时最重要的单元则在他们的全部回忆尝试中占统治地位。我们的数据和延多维斯卡娅(Yendovitskaya)(1971)及克赖斯蒂(Christie)和舒马赫(Schumacher)(1975)的报告是一致的；甚至学前儿童在他们详细复述故事时也喜欢抓中心思想而无视次要的、非必需的细节。

所以，虽然儿童的回忆确实反映了成年人的模式，即重要的单元比次要的单元回忆起来更容易，但是年龄较小的儿童仍然不能事先估计课文材料中各个单元的重要性。这就说明，8—12岁儿童在课文学习中预报单元难度时所体验到的问题和6—8岁儿童在词表学习任务中预报项目难度时所遇到的问题一样的对于决定一段文章的关键点有困难的儿童就很难期望他们能比较策略地选择这些关键点来精读。

4.关于策略活动结果的预报。 前一部分，我们讨论了儿童对任务难度的理解。这一部分我们要考虑儿童能否正确评价——他们自已的或别人的——记忆任务中所包含的策略活动。克罗泽尔等的问卷包含有这样一个题目，即学习时间。我们认为这个项目是一个很好的示例，可以说明儿童对策略使用的预报能力。给儿童20张彩色图片，告诉他们：有两个孩子看过并学过这些图片，一个学了5分钟，一个学了1分钟。要这些孩子们预报哪一个会记得比较多一点并说明理由。进而要求他们指出他们自已可能要学习多长时间，5分钟还是1分钟。正常儿童和智力迟钝儿童的数据见表2.9。

表2.9　　学习时间项目[a]：选择1或5分钟[b]的被试比例

被试	预报记忆最多的时间			适当的说明	被试的选择	
	N	5分	1分		5分	1分
（心理年龄）6岁						
正常幼儿园[c]儿童	20	0.75	0.25	0.35	0.65	0.35
首次参加实验的儿童	21	0.81	0.14	0.29	0.71	0.29
有经验的可教儿童	40	0.85	0.15	0.56	0.69	0.31
（心理年龄）8岁						
正常三年级儿童	20	1.00	0.00	1.00	0.95	0.05
首次参加实验的儿童	28	0.86	0.14	0.57	0.93	0.07
有经验的可教儿童	30	0.93	0.07	0.70	0.93	0.07

a.取自克罗泽尔等人的问卷(1975)

b.评定者间的信度＝0.99

c.正常儿童的资料取自克罗泽尔等人(1975)

大多数儿童都预报学习5分钟可能是较好的策略。但年纪较小的正常儿童和可教育的智力迟钝儿童不能说出正当的理由。此外，心理年龄6岁的儿童虽然也说时间较长是有好处的，但并非总能指出他们自己需要学习5分钟。

在我们实验室最近正在进行的研究中，我们考察了正常儿童和智力迟钝儿童，在对自由回忆学习的过程中使用策略的作用，能否正确评价。给儿童观看一段录像。录像所表现的是一个12岁的儿童在进行着四种不同的学习活动，这个儿童正在试图学习12张图片。所看到的这四种活动为分类、背诵、标记和注视。在儿童观看以后，要求他(她)指出哪种活动会有较好的结果。重复整个的预报程序以便获得可靠的分数。在第二次预报以后，给儿童同样

的一叠图片，告诉他用他所希望的任何方式进行学习以便学得尽可能多一些。

表明对某种活动有一致爱好的那些被试的预报结果见表2.10。14个正常的幼儿园儿童和11个心理年龄为6岁的可教育儿童和六个心理年龄为8岁的可教儿童的结果没有包括在内，因为他们的预报很不一致。首先我们看可教育的智力迟钝儿童，心理年龄为6岁和8岁的这两组都预报分类和背诵这两种合适的策略将导致较好的作业。没有人预言标记和注视这两种策略是合适的。然而正常的4岁儿童在他们的预报中对所有这四种活动相对地来说都是随机决定的。相反，心理年龄为6—8岁的可教儿童却能正确评价一个主动策略的价值，而实足年龄为4岁的幼儿却不能。但是到了三年级大多数正常儿童都能预报一个主动的策略是记忆所应采用的最好的活动。

表2.10 预报下述活动将得到最好结果的比例

活　动	分　类	背　诵	标　记	看
学前(实足年龄4岁)	0.24	0.33	0.13	0.28
一年级	0.44	0.25	0.09	0.22
三年级	0.35	0.46	0.19	0.00
可教的儿童				
(心理年龄6岁)	0.64	0.36	0.00	0.00
(心理年龄8岁)	0.38	0.62	0.00	0.00

注：引自"策略介入结果的预报"A.L.布朗，J.C.坎皮尼，C.R.巴克利，S.C.劳顿和R.琼斯，伊利诺斯大学，学习研究中心。

虽然我们将在以后再来考虑作业和预报之间的关系。不过为了作为这一部分的结束，在这里也要谈到在进行了预报后儿童的实际作业。关于分类和背诵这两种"合适"策略的数据被交叉折叠

在一起了。我们所关心的是那些能预报一个主动的策略是优越的和那些能采用一种策略的被试的比例。这些结果见表2.11。虽然可教儿童预报一个主动策略的优越性的能力是很感人的，但他们的实际作业较差。所有可教儿童都预报分类或背诵可能导致较好的作业。但是在看过录像后，面临着识别任务和即刻刺激时只有28%的儿童实际采用他们自己预报为优越的那一种活动。根据实际的作业，可教儿童和正常学前儿童及一年级学生没有区别。三年级学生倾向于采用他们自己预报是优越的那种策略。但即使是这些儿童，预报和作业之间的关系也是不准确的。最近我们正在从五年级的样本中获取数据。审查一下这些数据就可以看出即使五年级的儿童也不是照例采用他们预报是优越的那种主动策略的。

表2.11　预报分类或背诵好，并采用这些策略的被试比例

被　试	N	预报	策略活动	P(活动/预报)*
学前(实足年龄4岁)	46	0.58	0.13	0.22
一年级	32	0.69	0.25	0.36
三年级	26	0.81	0.62	0.77
可教儿童				
(心理年龄6岁)	14	1.00	0.36	0.36
(心理年龄8岁)	21	1.00	0.19	0.19

注：引自"策略介入结果的预报"A.L·布朗，J.C.坎皮尼，C.R.巴克利，S.C.劳顿和R·琼斯，伊利诺斯大学学习研究中心。

C.计　划

儿童事先计划的能力和他们对这种计划效率的知识是他的元记忆信息贮存的一部分。对这种知识的了解既有问卷式调查也有实验研究。我们先考虑问卷资料。

1.问卷资料：克罗泽尔等提出的那些题目中有几个直接涉及到记忆情境中的事先计划问题。我们选了其中两个被认为是特别适当的项目，即刻一延迟项目和学习一计划项目。在即刻一延迟项目中，问儿童在得知了他们朋友的电话号码之后，马上去打电话呢，还是先喝一点水？然后问他们如果必须记住一个电话号码的时候，他们该怎么办？总的回答结果见表2.12。三年级的正常儿童中有95%指出他们要先打电话，或者指出他们觉察到了迟延将会带来的问题。幼儿园儿童中有40%表现出了类似的觉察。可教儿童的结果在这两种水平之间。许多儿童指出如果要求他们记住的话，他们要写下电话号码。所有三年级儿童都表明有些计划性，但幼儿园儿童中有40%不能。一般地说，许多可教儿童以及正常的幼儿园儿童没有表现出记忆的计划性。

表2.12　即刻一延迟项目[a]：作出各种选择的被试比例[b]

被　试	N	先打电话或觉察到延迟的问题	记　忆　活　动			
			写下	背诵	其它	无
(心理年龄) 6 岁						
正常幼儿园儿童[c]	20	0.40	0.55	0.00	0.05	0.40
首次参加实验的儿童	21	0.65	0.30	0.00	0.00	0.70
有经验的儿童	40	0.46	0.46	0.00	0.00	0.54
(心理年龄) 8 岁						
正常三年级儿童	20	0.95	0.80	0.10	0.10	0.00
首次参加实验的儿童	28	0.75	0.96	0.00	0.04	0.00
有经验的儿童	30	0.67	0.63	0.00	0.03	0.23

a.引自克罗泽尔等人的问卷(1975)

b.评定者间的信度=0.93

c.正常儿童的数据来自克罗泽尔等人(1975)

表2.13　　学习一计划项目[a]：作出各种选择的被试比例[b]

被　　试	N	计划一种策略	未计划一种策略
(心理年龄)6岁			
正常幼儿园儿童[c]	20	0.45	0.55
首次参加实验的儿童	21	0.24	0.76
有经验的儿童	40	0.21	0.80
(心理年龄)8岁			
正常三年级儿童	20	0.90	0.10
首次参加实验的儿童	28	0.25	0.75
有经验的儿童	30	0.20	0.80

a.取自克罗泽尔等人的问卷(1975)

b.评定者间的信度=0.93

c.正常儿童的数据取自克罗泽尔等人(1975)

学习一计划项目是为测验儿童觉察出学习一套分类的图片的策略而设计的。在表2.13中，把儿童所选择的活动划分为策略反应和非策略反应。如果儿童指出他可能采用分类、联想、背诵或外部贮存等各种方法，那么这些都指定为策略反应，如果被试指出他要看一看或随机地排列项目，或者什么都不做，那么这些都是非策略反应。正常儿童和智力迟钝儿童完全不一样。到了三年级我们再次看到所有儿童都对这种任务表现出了一定的计划的行为。

2.分化假设。　为以后的回忆进行计划可包括属于学前儿童胜任能力的一些简单行为，但是这种计划也可能包含策略活动的复杂协调模式。首先我们讨论在典型的记忆范例这个范围内的简单的计划形式。幼儿能在他们必须设法去记住和不需要记忆这两

种情境之间加以区分吗?有关这种分化假设的早期研究提出,如果指令他们记忆图画与仅仅只是看图画,直到小学一年级的儿童才能作出不同的行为反应。可是所用的这种任务混淆了儿童对这种情境下如何记忆缺乏知识与他对记忆要求本身的觉察。如果主动介入的唯一尺度是在儿童的指令系统中还没有的策略的某种指标,这就很难想象我们怎么能指望儿童表现出他们对记忆指令意义的觉察。尽管儿童可能完全觉察出记忆和注视之间的差别,如果在学习期间还不会背诵、分类等等,那么就不可能期望从外部行为的表现看出他们是不要求记忆的儿童。

后来的研究进一步证实了这种判断。当把较长的注视时间作为分化的一种测量时,4岁半的儿童根据记忆的指令可做出不同的行为。然而甚至注视时间测量也是对所含的元记忆判断的复杂性非常敏感的。罗戈夫(Rogoff)纽科姆(Newcombe)和凯甘(Kag-an)(1974)告诉4、6、8岁的儿童,几分钟之后或一天或七天以后要测验他们对40张一套的图片的再认。只有8岁儿童能按照预定的延迟时间调节他们的审查时间;年龄较小的儿童没有表现出这种计划性。

不过,如果任务非常简单,在3岁儿童中也可以看到有这种计划行为的表现,韦尔曼里特(Ritter)和弗雷沃(1975)调查了一种非常简单的记忆情境,象阿普尔(Appel)等人所研究的那样,这种情境不依赖于机械记忆技能的掌握。托儿所的孩子看到一个实验者把一个玩具藏到许多相同的杯子中的一个的下面,实验者告诉其中一半的孩子说“和玩具一起呆在这里。”告诉其余的孩子说“记住玩具在哪里”。然后离开房间45秒钟。要求记住位置的那些儿童表现出努力设法记住这种信息。例如摸摸那个下面藏有玩具的杯子或者用某种方法使它能有所不同。关于“视觉背诵”的一个令人振奋的例子是有一个小女孩,她挨个指着这些杯子,对没藏

有玩具的杯子摇头，而对藏有玩具的那个杯子点头。要求记住位置的那个组比等在房间里的那个组回忆效果好一些；要知道这些孩子只有3岁。

在不要求机械记忆而要记住空间位置的情境下，幼儿确实对记忆意味着什么有点概念。识别客体的位置即使对蹒跚学步的儿童也是一个有意义的任务，但机械回忆一系列项目通常是进正规学校前的儿童所遇到的普遍的任务(布朗，1975)。所以不用奇怪，当儿童还缺乏把精细的记忆程序变成行动的先决技能时，他们自然不会从记忆的指令中有所获益。这一系列研究为诊断某一特定情境下缺乏元认知技能的问题，提供了一个告诫的例子。一个被试，儿童或成人，是否具有元认知洞察力取决于任务的难度水平，以及任务要求和被试现存的认知技能之间的匹配程度。

3.学习时间分配。　当面对着设法牢记一套材料这个普遍任务，而时间限制和其它限制又不允许从容地进行时，我们怎样来安排时间，以达到最佳效果呢？这个任务可能包含了象任何学生所能表明的，元记忆判断的非常精细的程度。由马苏尔(Masur)麦金太尔(Mcintyre)和弗雷沃（1977）引入了一个比较简单的实验类比。给一三年级的儿童和大学生被试同样提出一个多次尝试自由回忆任务。除了第一次尝试外，每次都允许被试只选择全部项目的一半来进一步学习。比较策略的作法是选择那些先前没有回忆过的项目来进一步学习。马苏尔等发现三年级学生和大学生确实选择先前遗漏了的项目进行额外的学习。但一年级学生就不是这样了。他们是随机选择的。作者总结道："把自己的学习活动有意地集中到还没有很好掌握的那一部分学习材料上的策略象其它基本记忆策略(如，机械背诵)一样，都不能机械地假定是幼儿学习技术指令系统的组成部分。

马苏尔等人的研究中一个有趣的特点是采用确认是"好"的策

略与有效学习的关系。虽然就大学生而言，策略的使用和作业之间的关系是相当清楚的。但策略使用的效果在三年级学生中还不是那么清楚的，对最年幼的儿童来说就更不清楚了。选择回忆过的项目和选择以前漏掉的项目，一年级的学生似乎同样能从中得到好处，即使假定后者是最佳策略的话。甚至是三年级学生也好象从选择漏掉的项目中只得到一点好处。在策略选择与良好作业之间缺乏联系的一种解释就是关于学习情境的要求。为了有效地作业，被试可能需要：1)识别出遗漏项目——马苏尔等人已经表明这对幼儿来说并不是困难的任务。2)选出这些项目来进行额外学习。3)在学习先前遗漏的项目的同时，大概还要通过背诵使以前回忆过的项目保持记忆。我们猜想问题就出在3)上。虽然幼儿可能再认遗漏的项目甚至可能觉察出他应该选这些项目来加强学习，但是，如果他能使以前回忆过的项目保持记忆的话，这个策略才是唯一有效的。对一年级学生来说，如果不背诵是困难的。对三年级来说也不是太满意的，这种三个要求的任务是不可能的，所以降低了策略选择的任何积极作用。

布朗和坎皮尼(1977)企图用可教育的智力迟钝儿童来重复马苏尔等的研究。研究的主要特征是它的训练性，在这个研究中对几种策略进行了考察。这里有趣的是预试的结果。没有一组被试表现出上述关于遗漏项目的偶然策略选择，但是，当考虑个别被试的数据时，几乎所有儿童中的$\frac{1}{3}$都选择了先前遗漏的项目。然而，这种选择并没有改善回忆的准确性。所以预试结果基本上证实了通过正常幼儿所发现的模式。大多数可教儿童不进行策略选择，甚至集中学习先前遗漏了的项目的那些儿童也没有从这种预见中得益。

策略的学习——时间分配涉及到那些不是机械学习一系列毫不相关的任务。在教育上特别感兴趣的就是学习课文。虽然我们正在

开始识别出成年人理解和记忆散文所用的有效策略，但对儿童的这些能力的发展我们却知道得很少。对有效地控制这种理解策略所需的自我觉察我们知道得就更少了。这决不意味着自发地使用各种元记忆技能来提高回忆是没有经过专门应用训练的中学生和大学生学习行为的可靠特点。然而，指导学生勾画出所学材料的轮廓并集中于主要事件，排除次要材料也是一种普遍的教育实践。正如我们已经看到的，儿童很难从一段复杂的散文中分离出主要的事件；所以他们不可能接受包含丰富信息的教学。

D.核对与监控

在预报和计划这两部分里，我们主要谈到了儿童在试图解决一个问题之前能首先考虑与任务有关的一些方面。在这部分，将要考察他们对学习或记忆结果的核对与监控能力。和前面那几部分一样，许多例子来自于有关问题解决的文献。因为在记忆研究中对核对与监控的研究还很少。在考虑这里所用到的材料时，我再一次感到震惊的是在相当大的一个年龄范围内都存在着这样一个"元认知的问题"，因为在核对他们作业的适当性时，并不是年龄小的孩子不如大人。相反如果被试没有充分地理解任务的性质，那么无论年龄多大都会表现出核对不足。

1.内部一致性。 核对不足的一个例子是儿童明显地甘愿接受矛盾的答案。在皮亚杰守恒任务的文献中内部不一致的例子是很多的，并且表明真正守恒的一个尺度是儿童能觉察出正确答案可以用各种方式加以核对并且这种核对的所有结果都应该一致，〔谢菲尔(Schaeffer)，埃格尔斯顿(Eggleston)和斯科特(Scott)(1974)〕。然而，与那些比较成熟的问题解决者相比，幼儿不大为他们自己的矛盾所困扰。我们作了一些关于重排顺序的研究。我们可以看一下其中的一个例子。让儿童重排以前看过的一个关于叙述故事

的图片顺序。要求他们从八张可能的图片中选出四张。八张当中有四张是原来故事的组成部分，有两张是用来分散注意的，画片显然是不正确的，而剩下的两张虽和事件的顺序相一致，但是它没有作为原始故事的一部分出现过。可以试做这个任务的学前儿童中，有41%选择了所有六张可能的画片并且能正确地排出顺序。但是他们不能指出六张画片中哪一张是他们自己以前曾经看过的。典型的对话如下：

实验者："这个故事中的哪一张图片是你看过的？"

被试："这些都是"（指着他们重新构成的故事中的六张图片）。

实验者："有多少？"

被试(数)："六张。"

实验者："你看过几张？"

被试："四张。"

实验者："哪些是你看过的——捡出四张来。"

被试："这些我都看过。"

这些幼儿完全没有对他们的矛盾答案感到不舒服。到了幼儿园的年龄这类结果就看不到了。

虽然我们有证据表明二年级的孩子确实对他们答案的内部一致性进行了核对，例如他们不会把一些意义矛盾的句子看成是同一个故事的组成部分，不过在某些情况下他们仍然可能接受这种不一致性。显而易见，学校算术问题提供的情境恰恰象霍尔特(1964)所给的例子数目将证实的那样。考虑下面的例子，给两个五年级学生一道题：256＋327，并且和教师一起逐步演算得出正确的答案583。清楚地记住这个和，然后教师在黑板上写出下一题，256＋328。同时指出在这两道题中他们都要在256上再加一个数，"我们现在是要加328而不是327。"希望孩子们能认识到这道题的

答案只要比上题的答案多1。不幸的是孩子们从头开始经过费力的演算才得出答案。教师又写出一道新题，并且一步一步地运算，直到孩子们相信了运算是正确的为止。立即擦掉，然后教师再把这一道题原样写出来，但是还得出错误的答案。现在在黑板并排写着两道题和其答案245＋179＝424、245＋179＝524。孩子们对这些解答都是非常满意的并且认为两道题都是对的，纵然单单根据内部一致性，他们也必须否定其中的一种解法。

2.现实性检验． 霍尔特举了一个有说服力的例子。儿童没有想到数学是"有意义"的．所以他们常常不能采用其它有用的核对手段，现实性检验或一种"注意错误，消除无意义的方法"就不足为怪了。确实在霍尔特的例子中有几个五年级学生在这一方面可以说是有病理缺陷的。霍尔特(1964)把一个孩子描述为情绪上以及在智力上都不具备核对其工作的能力，也不能把她的想法与现实相比较或者对她的思想的价值作出任何判断〔P48〕。霍尔特从初等数学课中举出的更为惊人的例子可能不会使大学的初等统计课的教师在遇到显然聪明的成人具有类似的符号休克症时感到震惊。大学生对现实性检验决不是没有偏见的；如果他们认为所用的公式是对的，就很容易把负的概念和方差做为解答。

如果儿童没有认识到一门功课，尤其是数学，应该有意义的话，那么就不会根据常数来核对答案。看下面这个五年级的学生。

> 一个男孩，相当不错的一个学生，正在做一道算术题。"假定有六个罐子，你要在每个罐子放$\frac{2}{3}$品脱的柠檬水，需要多少柠檬水？"他的答案是18品脱。我(霍尔特)说："每罐是多少？""$\frac{2}{3}$品脱"，我说"比一品脱多还是少？""少"，我说"有多少个罐？""6个"。我说："但是不合道理呀。"他耸耸肩并且说"那是按规律算的呀"。

现实性检验的一般概念怎样应用于这个问题领域呢？虽然在记忆文献中现实性检验的例子很少，这不是因为问题不存在，而是因为它还没有得到考察。然而，按照常识标准来核对某一种程序这个一般概念可以象应用于问题解决的任何其它形式那样也可以应用于记忆任务。在运用记忆策略时，儿童能通过提“背诵有意义吗？”“按照所需要的回忆类型有没有好处？”“所作的大量努力是不是合理？”这样一些问题来评价他们所用程序是否合适。这些研究所涉及的是儿童对记忆程序的觉察，而对记忆程序的觉察不是做为一种需要执行的过分学习方法，而是急需的一种有用的思维工具。

3.盲目规则仿效。 下面的这些例子也是从关于儿童数学问题解决的文献中选来的。所有例子中用到的都是各种特别设计的材料，是用来为核对结果、运算提供具体方法的材料。概要地用了乔仙尼亚(Cuisinaire)材料，概要地说，就是由长度从1—10cm变化不等的，宽1cm，高1cm的棒构成的。每种尺寸的棒有一种固定的颜色。各种颜色与长度的对应都由一年级小学生进行了充分的学习；这些学生后来用这些材料帮助他们自己掌握越来越复杂的运算——他们确实如此吗？

我们看一下爱德华(Edward)的例子，他是霍尔特的一个五年级学生，已经学过了计算规则。给他15很长10cm的棒和1根长4cm的棒，问他这些棒包含多少个厘米单位？(答案是154)。他把10cm的棒先排起来，把4cm的棒排在最后。然后他开始数排数并念到10、20、30……一直到100，数到这里时他开始摸着剩下的六很，念到200，300，400，500，600，604；他的答案是604。让他再试一次，他知道一定是什么地方错了。象以前那样第二次数到100还是对的，然后开始念到101，102，103，104……109，他的答案是109。别忘了15根10cm的棒的长度和颜色与爱德华已经用了好多年的长度

和颜色是完全一样的。这时尝试了一下教育干涉。材料被分成两半，第一部分为100(前10根10cm的棒)，第二部分为54(后5根10cm的棒和那根4cm的棒)。现在，爱德华毫不犹豫地回答："对了。"然后把这两部分在他面前合起来再问爱德华原来的那个问题，他用原来的程序做，又得604，爱德华至少是前后一致的。

然后引入了另一种训练措施，给爱德华100个棒（10个10cm的），问他有多少。一个容易照做的过度学习任务。然后把单独的白色单位(单个)加上去，一次加一个，当每个单位被放到它前一个单位的顶上时，爱德华能正确地数101，102，103，直到109。可是到最后一个放到顶上以便正好有11排，而每排10个单位的时候，爱德华却说是200。爱德华的问题在哪儿？他已经学会了当达到一个转折点的时候，要改变计数的单位。但是没有学会为什么改变和怎样改变。他一定是没有懂得这个任务是有意义的。

这种盲目规则学习的起源，在一年级开始使用这种棒的时候就很容易看到。每个儿童在一年级就都已经知道了1到10的每根棒的名字和颜色，并用这些棒来进行计算；他们能够数到一百，而且还能处理象"十"、"单位"这样一些概念。霍尔特表明，虽然他们在固定的情境下能够完成这些运算。但他们并没有理解棒所依赖的基本原则。他问一年级的学生"如果我们从桌边开始排38个白色的棒(1)排起来能排多远？"有一个孩子马上拿出了个桔黄棒(10)和一个褐色的棒(8)把它们排列起来。其余的孩子都试图把38个白的一个一个排列起来，在排的过程中常常还有几次忘记数到哪里了。这对于完全能背住桔黄色棒是10，但不理解它完全等于10个白色的单位的儿童来说是一个任意使用标记的极好例证。"6是暗绿色的棒碰巧有的名字，它的大小和其他棒没有关系"。

雷斯尼克和格拉塞(1976)也报告盲目规则仿效的一个引人注目的例子。教给5—6岁的孩子怎样用积木来得出矩形的面积。

然后要求他们得出平行四边形的面积。这是沃特海默（Wartheimer）平行四边形问题的变式，正确的解法是从顶角到底边作一条垂线，把垂线左边的部分移到图形的右边，这样就创造了一个矩形。雷斯尼克和格拉塞对这种创造性的解法没有发现什么证据，更有趣的是他们发现了儿童不能进行核对的十分显著的例子。许多儿童试图应用已经学过的规则，并且企图把这些积木放在平行四边形上，而丝毫不考虑右边角上缺一块。所以那些积木被挂在边上，但这些儿童在这样做的时候好象是完全没有困难似的。

霍尔特主张如果在训练儿童的问题解决规则和方法时，不同时使他们体会到规则背后的基本原理，就会直接导致象爱德华那样的盲目的问题解决程序。可以举这样一个例子，如果设法反复灌输有意识记策略就能得出很好的结果。已经有人提出了存在着这样一种危险，如果加强儿童机械学习策略方面的训练那么可能会使儿童看不到与要记的材料之间的较高水平相互制约的可能性。例如，如果训练一个儿童机械地背诵一列数字如492618或917342，他也打算去背诵234567，而没有认识到对这种有意义的一列数字就用不着背了。这有点象卢钦斯（Luchin）的水罐问题那种问题解决任务。由于熟练地掌握了适用于一系列问题的成功的复杂解法，而使被试对一个可用非常简单的解法解的问题使用了复杂的规则。所以，受过机械背诵训练的被试可能比没有事先受过背诵策略训练的被试更少注意和利用在数字集合中的多余性。

关于记忆任务中的盲目规则仿效我们有某些证据，但不是直接的并且是在回忆准备状态的研究中通过不太成熟的儿童得到的。对可教育的智力迟钝儿童进行了关于回忆准备任务的训练，这种任务类似于弗雷沃等人所设计的任务(1970)。在每次试验中给儿童呈现一系列图片(是他的广度的 $1\frac{1}{2}$ 倍)，并且要不断学习这些项目，直到他确信能够按顺序记住所有的图片为止。累加训练三

分之一的儿童学会背诵，训练另三分之一预报说出下一张图片，余下的儿童作为控制组，只指导他们标出一种不需要自我测验的活动。以后我们再来讨论训练的成绩和局限性。这里有趣的是比较年幼的儿童(心理年龄 6 岁)的行为。某一自我测验策略的训练足以导致他们策略生成的长期改善，进行背诵训练的那些儿童常常如此，但这不能导致他们监控能力的相应改善。训练二周后，年幼的儿童象所训练的那样能背诵和预报，却不能进行准确的回忆(或者近似准确的回忆)，这是一种记忆活动充分监控的尺度。我们对这一结果的解释是因为儿童仿效了"盲目规则"程序，让背，他们就背。但为什么这样做就能有助于他们满足准备状态任务的要求，这个原因他们并不清楚。

4.洞察力。　洞察力和盲目规则仿效正好相反，历来都是对问题解决及教育干涉感兴趣的心理学家所关心的问题。在年幼儿童的问题解决中，关于他们有洞察力的解法证据不多。而且这样的解法通常只是在还没有教给儿童某一种规则时发生。对教育家来说这个问题是很难对付的。霍尔特的五年级算术问题解决情境的例子中大部分是否定的。其中只有三个完整的关于洞察力的例子，前面提到过的一年级学生用代表10和 8 的棒去测量桌子。还有就是下面的例子：做2/4＋3/5这个题，一个孩子立即说结果一定是 1 或者大于 1 。"你要加上大于2/5的数才能得 1 ，可是2/4比2/5大所以结果一定大于 1 。"(霍尔特，1964)与此类似，另一个孩子立即认识到1/2＋1/3＝3/4不对。"不对，1/3与1/4不同。"我花了 1 — 2 秒的时间考虑她是什么意思。因为1/2＋1/4＝3/4，1/2＋1/3不可能等于3/4(霍尔特，1964)。

我们只知道两个实验的例证，而且这也是很少受到注意的一个领域。在刚才我们描述过的布朗和巴克利(Barclay)(1976)的回忆准备状态任务中，年龄较大的儿童(心理年龄 8 岁)确实有机智

地使用训练过的策略的表现。教给他们预报或背诵，他们不仅保持了训练的策略，而且他们判断回忆准备状态的能力也明显地改善了。(这些结果见图2.2，并将在下一部分讨论)在预报和背诵这两组中，年龄较大的被试都显著地改善了他们的作业——有一种改善在训练后大约保持了一年。这个例子就是在我们还没有教给儿童怎样使用策略来确定达到回忆准备状态时，他们能够机智地使用规则的一些间接的证明。如果只进行适当的任务训练，那么自我测验程序就足够了。在任务要求得到满足之前决定继续用那种活动这完全是儿童的责任。

机智地使用策略的一个比较直接的例子是由巴特菲尔德和贝尔蒙特(1977)提供的。他们主要论述的是在策略运用上的变化是任务难度的函数。涉及到最初选择一个策略的灵活性和当一个策略不再必要时就放弃它以及当这个策略又可使用而恢复时的效率。基本的程序是这样的，提出许多不同的项目来回忆，观察被试选择一种稳定的背诵策略所需的时间数。然后，事先不告诉被试，接连几次反复呈现一个项目表，个体不再需要对这些项目进行积极的工作了。最后，不声不响地再引入一个新表，并且个体必须再次开始使用他所选择的策略来处理信息。与年龄较大的儿童和成人相比，年龄较小的儿童在这几个方面所花的时间都较长：1)开始时选择一个策略，2)在不需要时放弃它，3)再次需要时恢复它。

巴特菲尔德和贝尔蒙特的研究(1977)为成人运用一个甚至很简单的机械学习技能时也动用越来越多的智慧和灵活性提供了一个极好的例子。就我们所知，在文献中关于机智地策略使用这种实验的例子还很少。多数现有的例子是轶事记录，例如我们关于在全新的任务判断中大学生很快地抛弃不适当的策略的报告（布朗，1973a)。在这个领域中需要系统研究儿童何时和何地应用某种

策略，并是否将为训练所影响。另外，我们还没有看到这样的例子，即去考察在两个或多个对抗的策略之间作出机智的选择，而这完全是和现实生活中的记忆情境极为类似的。

另一个被忽略的研究领域是被试可以把一个完全掌握的策略进行创造性的变化后应用到一个新的任务中去。虽然研究这类问题其困难是可以想象的。可我们还是感到惊奇，为什么对这个论题的研究会如此之少。布兰斯福德(Bransford)等人(1977)提出了一条有趣的途径可能是有用的。和记忆专家谈话所作的记录可能会把这些专家在学习过程中所采用的那些运算揭示出来。关于有效作业的这种洞察力可以用来指导不太高明的记忆者。

E. 训练研究

虽然花费了许多心机，努力设法给那些认为使用策略有帮助的人反复灌输一些具体的记忆策略。但是训练元记忆觉察的提法却还是一个新发展。除马克曼(1973)的一个研究外，直接针对作为一个记忆者的儿童的自我意识的主要训练企图，已经在我们的实验室实施了。而且主要对象是可教育的智力迟钝儿童(IQ=60—75)。阅读过前几部分的人至此可以看到，这些儿童甚至做简单的元记忆测验都会有很大的困难。他们在完成记忆任务中比较被动的一个原因是他们对自己作为学习过程的动因缺乏认识而造成的。所以发起一系列训练研究的主要原因是想要看看1)可教育儿童的元记忆是否可以改善；2)任何改善是否可能持久和概括化；3)任何改善是否将导致记忆策略一般应用的同时改善。

用学习缓慢儿童进行这种训练研究的基本道理，是从一次争论中得来的。这个争论涉及到为达到记忆技能有效的或持续的改善，训练特殊的记忆术有否作用〔布朗，1974，巴特菲尔德、万博尔德(Wanbold)和贝尔蒙特，1973〕。问题在于人们对迄今为止训

练普通记忆技能的尝试所获得的成功半信半疑。出现的一般情况是可教育的智力迟钝儿童对适当的训练做出反应，并且各种训练过的记忆技能在回忆作业中显示出满意的改善。而且，似乎是遵循精心设计的广泛训练，在一定的时间内可以看到这种经验效果的持续(布朗、坎皮尼和墨菲，1974)。不幸的是，很难找到对新情境概括化的证据(布朗，1974，坎皮尼和布朗，1974，1977)。概括化的问题不是一个新问题，尤其对训练智力迟钝者来说。美国和苏联心理学家都已经提出，训练迟钝儿童的一个主要困难是他们不能把所获得的信息和他们获得这种信息的特定形式区分开来〔希夫(Shif) 1969〕。最近大量的研究对训练的概括化问题已经提供了令人印象深刻的实验证明。

对所训练的记忆策略的广泛概括化缺少令人信服的证据，这一点表明不大可能从这样的练习中获得教益，因而也就导致了某些研究者提出这种观点，即训练的努力应该指向作业的一般决定因素而不是特殊的技能或策略。他们提出，与其只在某一特定方向进行训练尝试，不如把训练的尝试指向关于一般策略的知识的发展上获益更大。如果我们对有效地改善儿童各种类似任务的一般作业有兴趣的话，那么我们必须同时考虑训练(策略使用)的具体收效和一般的利益(改善关于记忆任务的知识)。

为了考察这一点，我们已经对智力迟钝儿童的元记忆知识进行了一系列的训练研究。情况与任何训练研究一样，无论是针对具体的策略还是有关记忆的一般知识，必须根据我们称作为保持和概括化的这两个标准来考虑训练的效果。作为成功训练的第一个指标，显然想表明训练效果在经过一段时间以后，还可以看得出来，这是非常必要的。因为有一定的证据表明，正在成长的孩子，当不再特别指导他们继续某一训练时，他们倾向于放弃它。成功训练的第二个并且是更为重要的指标，乃是对于新情境的概

括化。因为不能表现出广泛的迁移，任何训练方案的实际功效必定会成问题。

1.保持。

(a)回忆的准备。在我们实验室所进行的最初的训练研究中(布朗和巴克利，1976)考察了回忆准备的评价。这个训练程序的主要结果在前面部分已经描述过了(Ⅲ.D.3)。简单地说有两个自我测验策略，背诵或预报，对可教育儿童来说要么采用其中的一个策略加以训练，要么把他们分配到一个控制组，只让他们标出各个项目。这是一种不需要自我测验的程序。训练之后进行四种后测——一个及时的后测(训练后一天进行)和三个非及时的后测。在及时后测中指示他们继续采用训练过的策略，而非及时后测则在一天、两周左右、一年左右以后再进行。主要结果见图2.2。

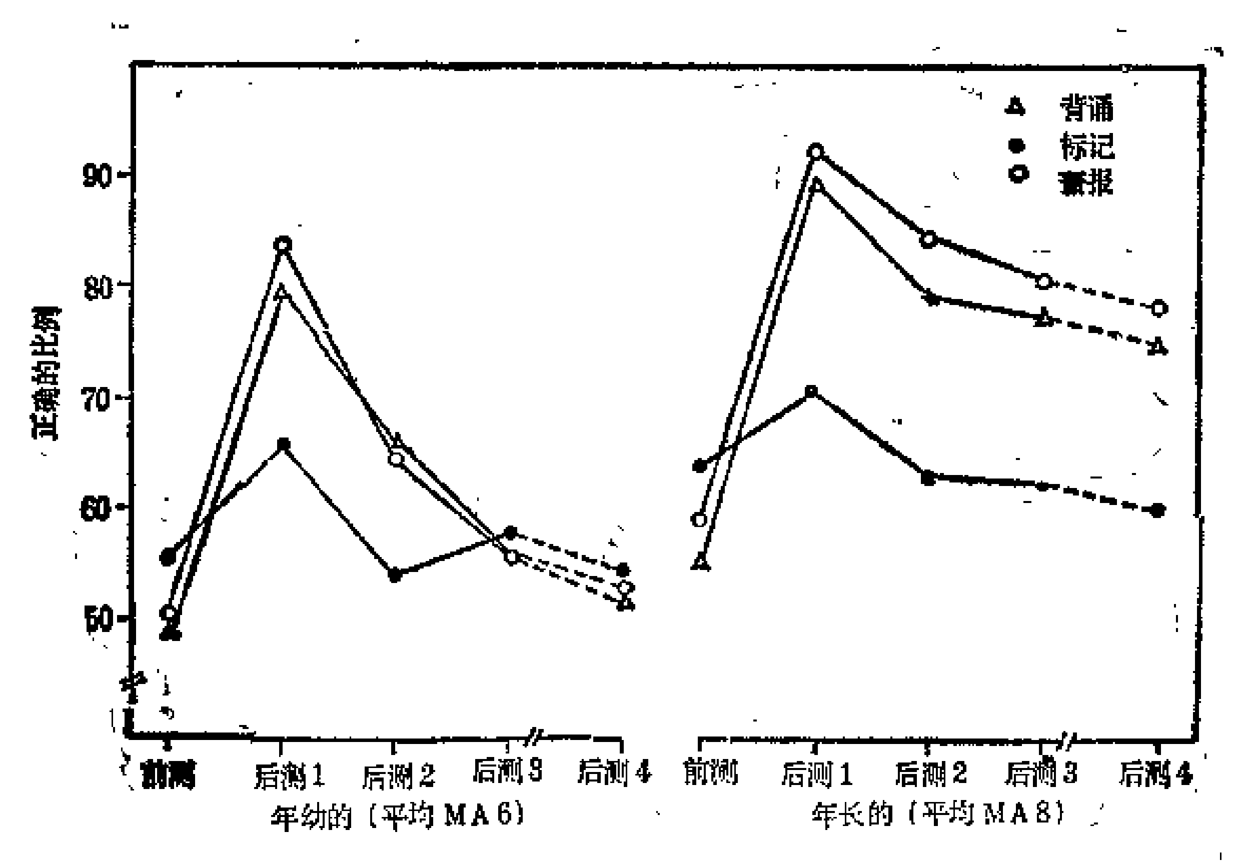


图2.2 作为心理年龄、训练条件和测验阶段函数的正确回忆比例

图中给了正确回忆的百分比。如图所示，(在背诵和预报组里)

无论是年龄小的或是年龄大的儿童及时后测(后测1)比前测明显得多。另外，在预报组和背诵中，18个年幼的被试中有13个至少在一次试验中作出了准确的回忆，与此相比在前测中没有一个人能做到这一点；年龄较大的被试在后测1中26个人有24个至少在一次试验中作出了准确的回忆，而在前测中只有2人。所以训练有用的自我测验策略既促进了作业(回忆数据的百分比)也改善了监控能力(准确回忆的数目)，值得注意的是标记组(控制组)没有表现出这种改善。

心理年龄6岁和心理年龄8岁的儿童在后三种后测(非及时的)中明显不同。年幼组在后测2、3、4中的作业和训练前的水平没有显著的差异，而背诵和预测组中年龄较大的儿童在所有各种后测中其作业都比训练前的水平有明显的差别。所以，和以前有关一种策略的直接训练的研究一样，训练促进了作业，其影响对年龄较大的儿童比较持久，而对年幼的儿童则是暂时的①。

(b)学习时间分配。在我们的下一项训练研究中，我们考虑策略的学习时间分配。这项研究的前测数据已经在前面讨论了(Ⅲ、c、3)。在前测期间，除了第一次是多重尝试外，其余的每次都是自由回忆程序，允许可教育的智力迟钝被试选择一半需要记忆的项目(6/12)，看他们是否会策略地选择那些遗漏了的项目来进行额外学习。进行了前测后，被试分成三组来训练，由实验者来为儿童选择学习项目。对第一组儿童(标准策略)，实验者是根据成熟诊断的策略来选择项目的，也就是实验者把儿童在以前自由回忆的尝试中遗漏的项目返回给儿童。对第二组采用了另一类系统的选择。实验者把被试已经回忆出的项目再加上一个新的项目返回给他(渐进策略)。这样做的道理是，如果还未达到成熟的被试因

① 初步的数据表明，对散文学习情境的概括化而引起了回忆准备的成功保持。这项研究正在进行中(布朗和坎皮尼未出版)。

不能保持住从前回忆出的项目而不能从关于遗漏项目的额外学习时间里获益的话，那么标准策略对他们的作用就值得怀疑了。渐进的程序能够使他们在每次尝试中只增加一项额外项目，同时允许他们继续复习先前回忆过的项目，所以他们能够逐渐地提高到较好的作业水平。在每次尝试中，第三组被试所接受的复习项目是随机选择的。在训练之后，儿童所接受的后测象在前测中一样仍然是自由的。他们可以选择愿意学习的那些项目，但条件是不能超过六项。

在前测和后测中要考虑平均正确率和标准化选择记分(分数越高,选以前没有回忆过的项目来进一步学习的倾向越大,细节可参考布朗和坎皮尼（1977).前测和后测回忆的平均正确率见图2.3,在此可以看到对于年龄较小的被试来说，作为训练条件的一个函数在前测和后测之间似乎没有什么变化。对年龄较大的被试来

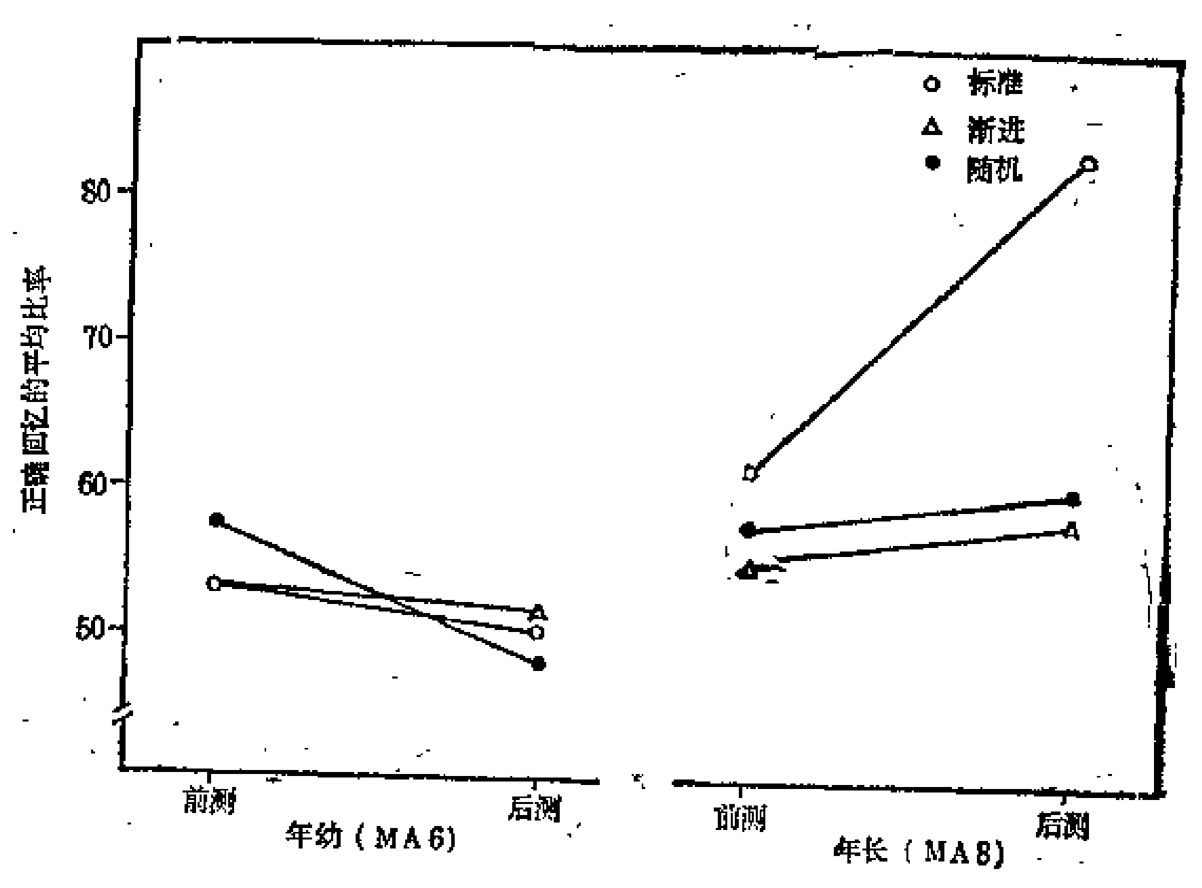


图2.3　作为年龄和训练条件函数的前测、后测的正确回忆平均比例

说，在训练标准程序期间，迫使学习遗漏项目的被试，他们的回忆比例在后测中显著地增加了。其它二种条件在前测和后测之间似乎看不出什么变化。所以只有在训练中被迫学习遗漏项目的年龄较大儿童其后测才表现出有所改善。平均标准选择分数和与前测比较的分数一起见图2.4。

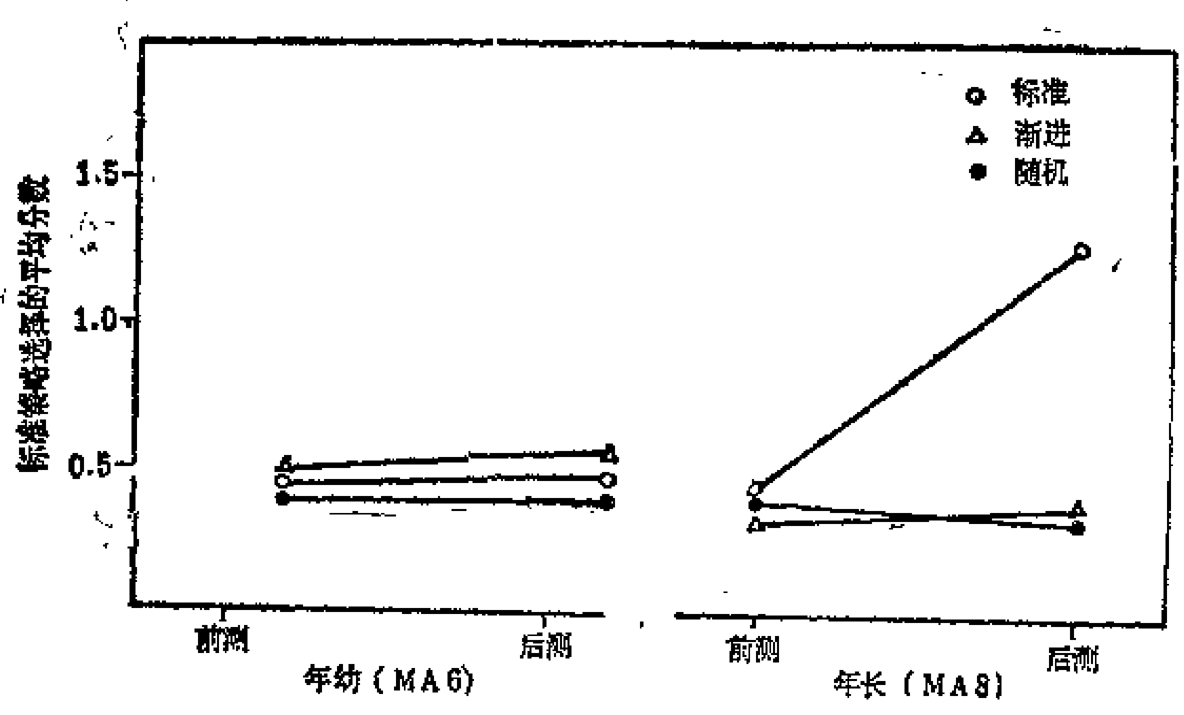


图2.4 作为年龄和训练条件函数的前测、后测中标准策略选择的平均分数

这里我们所看到的趋势如同在回忆分数中看到的一样。只有在标准条件下进行训练的年龄较大儿童其前测和后测之间才表现出有些变化。

除一组外，在其余各组被试中没有看到前测和后测之间的差别，这是训练的失败，还是迁移的失败？为回答这个问题，我们必须参考表2.15中呈现的训练数据。只有训练的回忆数据是可用的，因为实验者选择了这些项目来学习。在渐进的条件下，年幼的被试在交叉尝试中和分块尝试中都确实有所改善，但在其它两种条件下看不出什么改善。年龄较大的被试在标准条件下其交叉

尝试和分块尝试都确实有所改善。但在随机条件下没有表现出改善。然而，值得注意的是在渐进条件下年龄较大的被试在交叉尝

表2.15　　作为年龄和训练条件函数的训练回忆平均比例

被试	年幼的(心理年龄6岁)			年长的(心理年龄8岁)		
分块尝试[a]	1	2		1	2	
训练条件						
渐进的	0.36	0.46		0.46	0.44	
标准的	0.34	0.33		0.50	0.59	
随机的	0.29	0.28		0.41	0.42	
尝试[b]	1	2	3	1	2	3
训练条件						
渐进的	0.34	0.40	0.48	0.41	0.43	0.51
标准的	0.32	0.33	0.35	0.46	0.55	0.62
随机的	0.27	0.28	0.31	0.41	0.40	0.43

注：引自"可教的迟钝儿童的策略的学习时间分配"A.L.布朗和J.C.坎皮尼，Intelligence，1977，1，94—107。

a.儿童每天接受4个表，这些被分成两块来尝试。

b.对每一个表被试回忆四次。后三次是在项目选择后进行的尝试。

试中(当然不是分块尝试的交叉)有某些改善。无论是年龄较小还是年龄较大的儿童在训练阶段中都发现有所改善，所以年龄较小的儿童在后测中没有什么改进可以归结为是迁移的失败。

现在来对训练结果作个小结，幼儿可以从所给的渐进策略中得益，但不能从所给的标准策略中得益。而较大的儿童则从所给的标准策略中得益最多，从渐进策略或随机选择中得益甚少。这

个结果模式似乎进一步证明：所使用的策略要想获得成功必须和儿童的认知胜任能力相适合。强迫儿童按照成人的策略学习只能有助于那些可以满足(在一定程度上)那种策略特有要求的年龄较大的儿童。

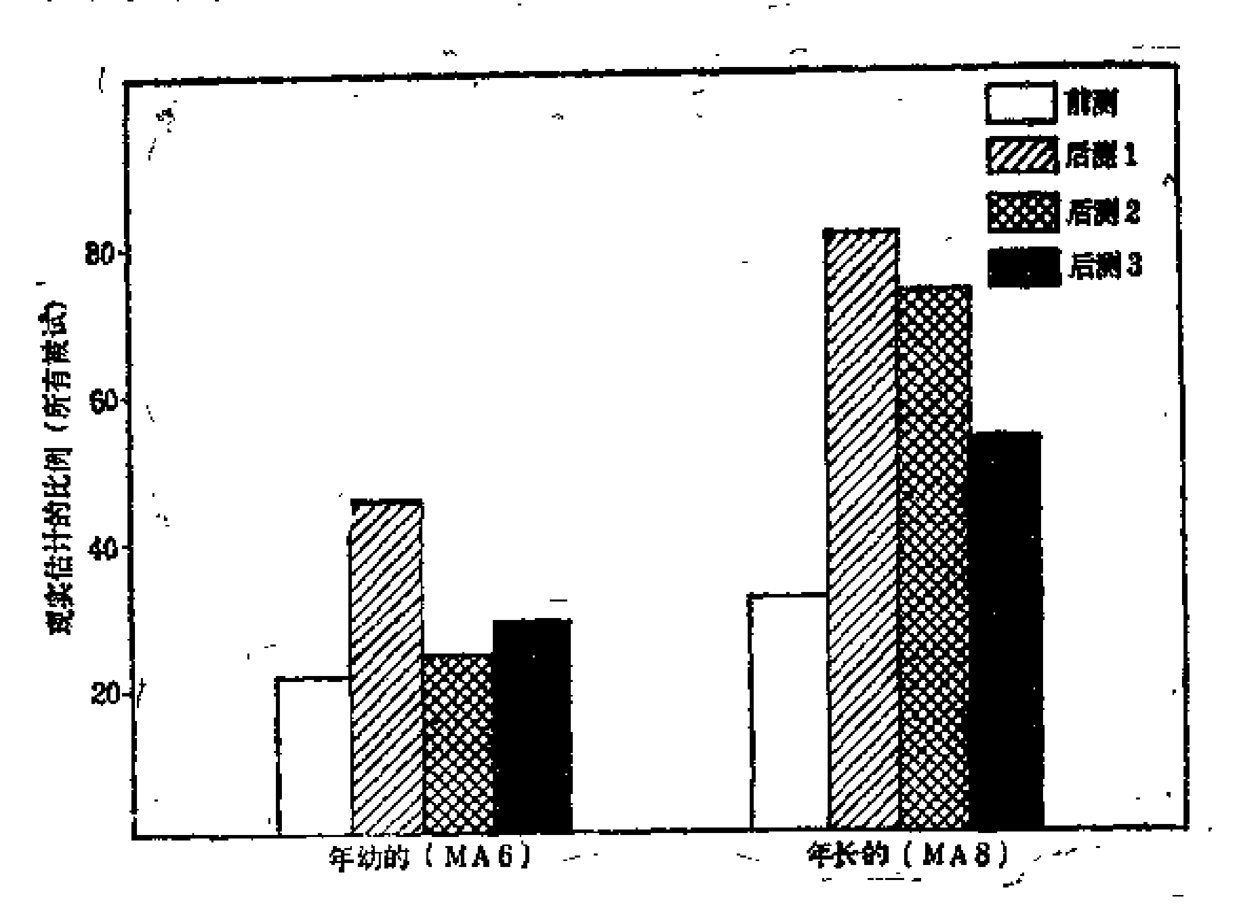


图2.5 作为心理年龄和测验阶段函数的现实估计的比例

(c)广度估计。 在这一点上，我们认为至少对那些不成熟的被试来说，一个好的研究策略主要是直接进行元记忆行为的训练，而不应该是布朗和巴克利(1976)以及布朗和坎皮尼所采用的那种间接的方法。还有一点变化是我们的研究重点不再是策略使用的监控，而这正是布朗和巴克利以及巴特菲尔德和贝尔蒙特研究的课题。运用与任务有关的记忆术及监控它的成败好象都涉及到内省与外显行为的复杂协调。无论正常儿童还是迟钝儿童这种协调都是发展得比较晚的。根据以往失败的教训，我们决定考虑一种

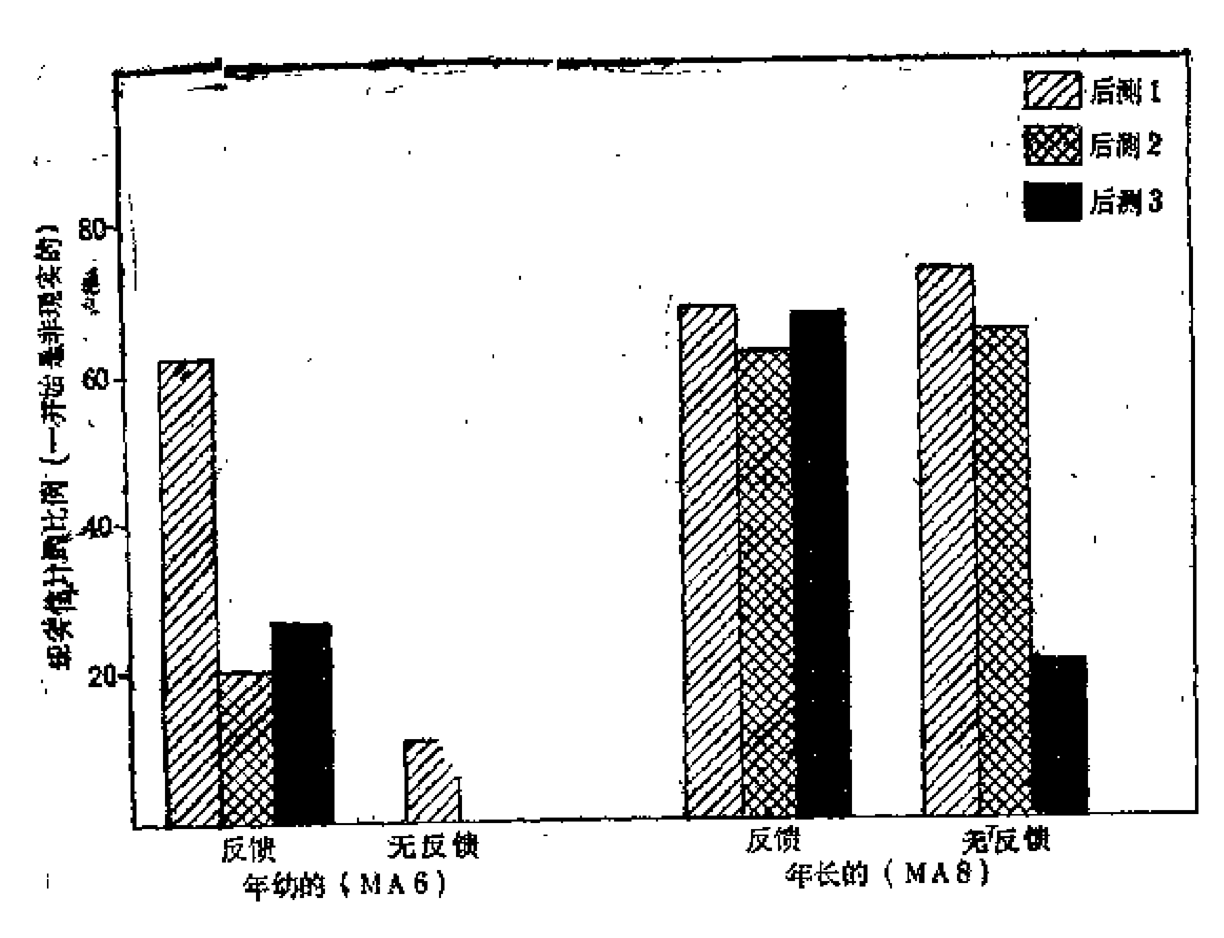


图2.6 作为心理年龄，反馈条件和测验阶段函数的现实估计的比例(只考虑一开始是不现实的被试)。

较为简单的元记忆觉察形式，也就是善于估计他们自己的广度的能力，这种广度估计的能力似乎为今后试图采用和控制某种特殊策略打下了基础。如果儿童没有觉察到他们自己的记忆限度有多大的话，几乎不可能期望他们能采取某些步骤来补偿他们的缺点。

之所以选择广度估计任务，还因为它一直是先前的训练试验的课题，并且所得结果稍有矛盾。马克曼(1973)发现62%的幼儿园儿童对10次明确的训练试验做出了反应，相反尤森和莱维(1975)却发现学前儿童对于表现出回忆不充分性的练习尝试所进行的反馈明显无动于衷。这些差别可能是由于年龄和明确训练的范围所引起的。因为我们通常需要用明确而广泛的训练来改善迟钝儿童关于特定记忆技能的使用。所以决定提供一种关于元记忆任务的

广泛而明确的训练。简而言之，我们希望提供一种最佳的训练经验，以便评价这样的训练是否可能导致年幼儿童对于自己的记忆局限性的认识的长期改善。

两组首次参加实验的可教育儿童(心理年龄 6 岁和 8 岁)，给他们看几套排列整齐的图片，每套包括10张图片(同时呈现)，让他们预报他们可能回忆出多少。〔在整个研究中，一半图片分了类，另一半没分类。这样做的目的是当要求学生预报其回忆的数目时，不至于记住一个具体的数(比如，6)。一个适当的反应应该表明对于有组织的、分了类的那些图片回忆起来的比较多。而且实验结果确实如此。〕。然后把儿童的预报和他们(后来测定)的实际回忆结果相比较。如果某一被试的估计值和他实际回忆的数相差在2项以内(±2)则我们称之为现实的估计者；那些猜测误差超过 2 项以上者称为不现实的估计者。只有31%年龄较大的儿童和21%年龄较小的儿童能够被归为现实的估计者。其余的人都过高地估计了他们的作业水平。(大多数人都预报他们能够回忆出所有的10张图片。)

所有的儿童随后都被给以连续10次训练试验。在这些试验中，要求他们事先估计自己的作业，然后回忆。在每一心理年龄组中有一半被试可以得到明确的反馈，对他们的预报给予提示，并且用视觉显示和口头语言报告的方式向他们表明他们实际回忆出的项目数。在每次估计和回忆之后都提供这种反馈。另一半儿童也进行预报和回忆，但是不提供这种明确的反馈。训练完成之后，进行三次后测。第一次是在训练后的第一天，第二次是在训练后两周，第三次是在首次后测的大约一年之后。

图2.5、2.6给出了我们极为关心的数据。图2.5给出了两个心理年龄组的被试在前测和每次后测中现实估计者的比例。一般来说年龄较小的儿童在第一次后测(训练后的第一天)中表现出某种

改善，但是两周以后就回到了原来的水平。而年龄较大的儿童最初的改善是很显著的，而且经过一段时间仍然维持得很好。甚至一年之后，现实估计者的比例(56%)明显地高于训练前的比例(31%)。

图2.6中只考虑了最初是非现实估计者的数据，接下来根据心理年龄水平和反馈变量来分析这些结果。一开始就成为现实估计者的学生在整个实验中始终如此。很幸运，我们的训练没有使得他们退步。从那些开始就是非现实估计者的儿童的首次后测看，年龄较大的受训儿童中有65%不依靠反馈条件就能变得现实了。而年龄较小的受训者中，给予反馈时有62%变得现实，不给予反馈时只有9%变得现实。从第二次后测的数据看，年龄较大的受训者的情况依然如此，有60%仍然是现实的，并且没有给予反馈。然而，对于年龄较小的受训者来说，给以反馈时有18%仍然是现实的。不给反馈时没有一个受训者是现实的了。所以从年龄较大儿童的前两次后测看，不管有无明确的反馈，训练本身就足以导致现实的估计并且效果是持久的。而年龄较小儿童所得到的结果则截然相反。只有在训练期间提供明确的反馈时第一次后测才有明显的改善；甚至在这种情况下，效果也不持久，现实估计者的比例从第一次后测的62%降到第二次后测的18%，对于年龄较大的儿童给予明确反馈的效果只是在最后一次后测中才看到。在反馈条件下现实估计者的比例仍然没有变化。相反对于那些在训练期间没有给以反馈的被试，只有20%大约到训练后一年仍保持着现实的估计。

(d)总结。这些初步的实验结果表明轻微智力迟钝的儿童在估计他们自己的作业上存在着问题。无论是执行任务之前还是执行任务之中都是如此。另外还有一点似乎也是清楚的，即对于年龄较小的儿童在看到任何改善效果之前，关于他们的作业必须给予

明确的信息，并且为了维持作业，需要不断地给予提示和督促。在所有三个实验中，关于训练效果的持久性可以看出有一个明显的发展趋势。年龄较大的儿童，训练有较持久的效果，而对年龄较小的儿童来说训练效果持续时间特别短。

2.概括化。 我们力图发现训练效果的稳定性，但结果不太理想，这就挫伤了我们对于概括化测验的热情。然而，正因在所有的研究中年龄较大的儿童表现出了适当的稳定性，所以我们也涉及到了布朗等人的特殊概括化测验。并且我们一般都看看年龄较大的儿童回忆准备状态的概括化。除了特种技能概括化的测验之外，我们也考虑了一种更一般的迁移现象。也就是说，已经经历过几次记忆和元记忆训练研究的儿童和没有受过训练的首次参加实验儿童之间，在而对每项新任务表现出的一般改善上或通过问卷法研究一般的元记忆觉察时，能看出有什么差别吗？

在继续描述我们的一般和特殊的迁移数据之前，我们要弄清楚关于概括化的标准是什么。我们把许多自称是进行概括化的研究看作是测量稳定性。因为他们只对概括化任务使用了新的刺激表。而我们在整个研究中都使用了新的刺激，并且假定对新的刺激表继续进行已训练过的活动就是关于稳定性的测验。概括化测验不仅涉及到新的刺激而且还包括其它某种变化。但是这种变化的成分是极小的。后而我们还会谈到概括化的适当标准这个问题(Ⅳ.F)。

(a)特殊的概括化。在我们所进行的关于训练概括化的特殊测验中唯一完成了的研究是布朗等人的(1977)广度估计训练。为了进行前测后测和训练测验，我们使用了一个关于广度估计的修正版本。也就是说，每次试验，让被试看上有10个小图片的大卡片，然后要他们估计可能回忆出多少。在前测和后测中，还包括了弗雷沃等(1970)所使用的顺次排列的任务。该任务的项目按增量的

顺序(1,2,3 直到10)呈现，而且在每次呈现时儿童必须指出他们是否能回忆出那个大卡片上含有的小图片。对于顺次排列任务现实估计者的比例两个组都是很低的，并且受测验时间的影响不大。(年龄较大的被试在前测和三次后测中是 18、18、20、15，而年龄较小的儿童则为03、07、03、05)，即使有一点改善，特别是在年龄较大的儿童中作为10个项目任务的训练结果，这种改善没有概括化到非常类似的顺次排列的测验中去。

在第二次后测之后，第一天进行了另外一个概括化测验。向被试呈现20张具有10个项目的卡片。每张卡片都含有1到10的数字。其中10张卡片所包含的数字是按序排列的，从某一个数开始而不是从1开始；其余的10张卡片所包含的数字是随机排列的。被试看过20张卡片后指出对每一张他都能回忆出多少。然后评价他对于两种材料实际回忆的情况。最后，把这些卡片配对，一张随机的，一张有序的，再问被试哪一种比较容易回忆，为什么。采用了这样两种排法的卡片：有序排列的和打乱顺序的。对有序排列的材料(如按序排列的数字)能预报出10个项目是一种现实的估计，相反对于打乱顺序的材料做这样的预报将是不现实的。因此我们分别考虑这两种类型的材料。表2.16中只列出了随机排列的材料的数据。首先来看看一开始就是非现实的被试。显然，没有证据表明在高度类似的10个项目的图片任务训练之后产生了概括化。能做出现实估计的被试的比例无论对于年龄较小的还是年龄较大的被试都是低的，而猜中10项的人数相当高。下面再看看一开始就是现实的被试(在各种反馈的条件下倒退下来)，这里情况就完全不同了，无论是年龄小的还是年龄大的被试其平均差异分数(预报的与实际的)都在现实的区域之内(±2)。几乎有三分之二一开始就是现实的被试对于数字概括化测验仍是现实的，而猜中10项的人极少。

回过来再看看关于有序材料的预报，类似的情景出现了。能准确预报他们将能回忆出 9 或10项(如评价这些材料的组织)的被试的比例对一开始就是现实的年幼被试和年长被试来说分别为0.67和0.58。一开始就是非现实的年幼被试没有人能做到这点，而年龄较大的被试只有0.26的人能预报出9或10个项目。

表2.16　　数字概括化测验中现实判断的比例

N	一开始是现实的		一开始是非现实的			
			无反馈		反　馈	
	年幼的	年长的	年幼的	年长的	年幼的	年长的
	(MA6)	(MA8)	(MA6)	(MA8)	(MA6)	(MA8)
	6	12	11	14	12	12
平均差异分数	2.83[a]	1.08	5.09	3.64	5.00	4.42
现实的比例	0.67	0.75	0.09	0.28	0.08	0.25
猜中10项者的比例	0.17	0.17	0.73	0.43	0.67	0.50

注：选自"可教的迟钝儿童训练后元认知觉察的稳定性和概括化"A.L.Bronn,J.C.Campione和M.D.Murphy，实验儿童心理学杂志1977，24，191—211。
a.这个数字可能有误。这一组只有6个被试，4个现实的和2个非现实的。4个现实的被试的平均差异分数为1.75，其余2个被试均过高地估计了他们的广度，相差5个单位。

最后，给儿童10个强迫选择试验，在这一试验中一张有序项目的卡片和一张随机项目的卡片相配对，并问他们哪一张较容易

回忆。表2.17表明了说有序的数字材料比随机的数字材料容易记忆的(即，在19次试验中有 9 次预报有序的强于随机的)被试的数目。又一次看到，几乎2/3一开始就是现实的被试预报说有序数字的卡片较容易回忆，而经过训练的现实的被试与非现实的被试看来没有多大差别。

这样一来，我们为发现特殊训练的概括化所做的一系列努力并不令人鼓舞。因为没有什么证据表明概括化是训练的结果。那些对于训练任务一开始就是现实的被试，的确表现出对所有概括化任务的迁移。这表明这些任务本身对于有能力的被试来说就是一些适当的迁移测验；然而受过训练的现实的被试则没有这样灵活。应该注意：这些概括化任务是和训练任务非常类似的。总之基本要求是估计自已关于10项表的广度。顺次排列法包括同一类的刺激，但是任务项目安排的形式稍有变化。在数字估计问题中，任务项目安排的形式(10项)保持不变，但刺激类型改变了。这两方面的变化都是很小的，而且都已表明对于提供一种理想的概括化的情境是极好的第一步。然而在年龄较大的被试中还没有发现什么。可能这个训练计划的惨败是由于这样一个事实，训练只是按一种任务进行的并且所获得的信息又确实是和这个特定的任务紧密联系在一起的。如果这一点是真的，那么下一步必须要进行各种类似任务的训练。所有任务都需要同样的策略，然后寻求对于新任务的概括化，这些新任务是和训练任务属于同一种类的。没能发现概括化的第二个原因是没有给儿童以概括化的训练或明确的提示(看Ⅳ.F)。无论如何，在探讨迟钝儿童认知训练概括化的难以捉摸的证据方面我们需要付出大量的时间和努力。

表2.17　　数字概括化测验中关于回忆容易度的判断[a]

	一开始是现实的		训练后现实的		非现实的	
	年幼的 (MA6)	年长的 (MA8)	年幼的 (MA6)	年长的 (MA8)	年幼的 (MA6)	年长的 (MA8)
有序的较容易	0.66	0.75	0.14	0.26	0.19	0.12
无序的较容易	0.00	0.00	0.14	0.11	0.13	0.12
不一致的	0.33	0.25	0.71	0.63	0.67	0.75

a表中的数字是属于每一类的被试的比例。

(b)一般迁移。我们原打算不厌其烦地去发现训练的概括化，看来这些努力没有获得鼓舞人心的结果。我们对有经验的被试和首次参加实验的被试做了比较，发现可教育儿童不能轻易地表现出概括化。首先我们在有经验的被试和首次参加一个新实验的被试之间没有发现可靠的差异。无论什么训练效果肯定不会强到足以损害我们进一步研究的被试总体，也不会强到可教育他们。

在力图系统地考察概括化时，我们运用了克罗泽尔等人(1975)的一般元记忆觉察的问卷。我们对来自同一学区的四组儿童实施了这种问卷。这四组儿童的情况是：心理年龄 8 岁首次参加实验的(N=28)，心理年龄 8 岁的有经验的(N=30)，心理年龄 6 岁首次参加实验的(N=21)和心理年龄 6 岁有经验的(N=40)。有经验的儿童至少参加过两次元记忆训练研究或者有时候是其它的问题解决和记忆实验。首次参加实验的儿童据我们所知从来没有作过实验被试。我们的想法是看看有实验经验的儿童是否比首次参加实验的儿童在各种问卷项目上都表现出更多的一般觉察。

在这些数据中有些已经给出了(表2.2，2.3，2.4，2.9，2.12和2.13)，细心的读者将会注意到明显缺乏经验的效果。没有表现出一致的情况，有经验的被试没有表现出比首次参加实验的被试有

更多的觉察，甚至是在那些与他们所接受过的特殊训练有关的科目（即接受过复述训练的儿童没有表明复述是可以采用的一种合理的活动或者说将导致较好的作业）。但有一点证据表明有经验的儿童对他们的反应更能作出适当的解释，而且也能作出更多的解释。这被认为是在与以前对他们做过实验的实验者进行谈话感到比较熟悉和无拘束的结果。在整个研究中出现的唯一可信的差别是在首项上。就这一项问儿童是否善于记忆，许多有经验的被试认为他们不善于记忆（心理年龄6岁首次参加实验的＝0.24，有经验的＝0.46；心理年龄8岁首次参加实验的＝0.46，有经验的＝0.64）。所以，两年训练的唯一可见的效果只是提醒孩子注意他们自己记忆的不足之处、而没注意到克服这些缺点的方法，比我们所希望的主动克服显得逊色。

Ⅳ. 元记忆：新思想和老问题

关于儿童元记忆的富有见解性的研究刚刚问世时，反应非常强烈，大家认为确实有了一种比较聪明的方法来研究记忆的发展：这种研究的出现表明重点明显地转移了，从注重儿童的机械学习技能转移到把儿童看作一个主动求知的积极分子，——而且，还是一个受着至今仍未了解的多种因素影响的积极分子。不再把记忆技能的发展和有关记忆的知识视为彼此分离的现象，而把它们看作是一定文化背景中的儿童的认知发展的一个组成部分。虽然我们认为这种重点的转移是一种令人振奋并卓有成效的发展，然而对这一领域将要采取的方向我们仍然有一些疑虑。元记忆发展的研究既然已逾五年，我们似乎有理由强调一下这些保留的意见并考虑这一领域的一些共同问题。我们愿意指出，我们自己的研究工作所受到的批评如果不比任何其它的研究工作所受到的批评

多，至少也是相等的。

A.演示研究的不足

关于记忆不同侧面发展研究的历史，是比较短的，直到20世纪60年代，这方面的工作才成为有关儿童研究主流的一部分。从60年代到70年代，我们亲眼看到，对以年幼儿童运用积极学习策略的能力为研究中心，越来越感兴趣。这些研究的大多数都涉及机械记忆技能，特别是涉及归类和复述的技能，这些技能都是回忆各项孤立材料的保证。对这个课题兴趣的高涨实际上产生了数以百计的研究。这些研究表明：一些发育中的年幼儿童：(1)用这些技能的能力不强；(2)运用这些技能的能力容易训练；(3)除非有人明确促使他们继续运用这些训练有素的技能，他们是倾向于放弃这些技能的。

在20世纪60年代末和70年代初，第一批元记忆研究问世。这些研究主要侧重于对儿童的基本机械记忆技能(事实上就是复述与归类这两种技能)的了解。这里，我们又可以看到历史的重演，因为我们发现越来越多的演示实验表明不善于运用基本记忆术的人少于熟练运用策略的人。我们悲观地看到，在今后的几年中能有同样大量派生的演示实验来把问题说得一清二楚。过去我们在这方面曾作过贡献，将来我们还会作出贡献。

虽然演示研究有助于增进了解幼稚的思维者如何不善于记忆，而且，这一领域的最初研究对人们了解这一问题至关重要，然而，我们对于这种演示研究的大量增加是否有益仍持怀疑态度。是否还确实需要更多的研究来表明幼儿并不是经常考虑如何思维、记住如何记忆，或者更没有学会如何学习的呢？真正需要的则是发展一种理论。这种理论可以使我们的演示研究更明智，更适应实际的需要，能指导我们在试图了解思维发展的过程中走捷径而不是

走弯路。我们知道这样一种理论的发展并不是一蹴而就的。各种如此的发展理论一般来说有几个特有的弱点，例如：这些理论要么太概括(这里没有反对的意思)，要么太专门，概括的范围有局限性；这些理论往往或者注重发展的结果，这实际上则忽略了作为发展主体的过程；或者注重过程而对发展中的棘手问题则不闻不问。所以对于理论的要求并不是一个无足轻重的问题。

在某种更为实用的研究计划中，为了更深入细致地分析所运用的具体范例，我们不主张进行孤立的演示实验。这里，我们再一次表示不赞成进行大量的范例研究；那些有损于详细分析所用任务的例子在文献中比比皆是。然而，似乎需要一个中间的状态。例如，只有两种元记忆范例研究了不止一次，即弗雷沃等人(1970)最初介绍的回忆的准备状态和广度估计这两个课题。这两个课题在程序上都有一个有趣的缺点，这样的缺点能导致对儿童元认知技能的过高评价。在回忆的准备状态任务中表的长度的增加超过了广度则会严重影响作业(布朗和巴克利，1976；马克曼，1973)。在广度估计任务中，在超过儿童能力的第一个指标处停止的作法也还可以对幼儿的元记忆作出过于乐观的描述(布朗等人，1977)。对任务要求缺乏深入的理解，不加批判地接受一个孤立的演示研究实在是一种危险的游戏。

B.元记忆还是元机械学习？

元记忆的研究不但在很大程度上局限于孤立的演示研究，而且几乎只注重研究儿童的知识，以及对为数不多的简单的机械学习技能的掌握情况。弗雷沃关心外部和内部行为的现实主义的探讨则是一个值得注意的例外。现在，大家对于过分地注重再现回忆的机械学习技能，特别是对于那些无意义的，孤立的材料表的机械学习技能感到痛惜，而对于元记忆的文献，人们则很少提出

这样的批评。儿童知道他能机械地学习实验室的材料，这并不是元记忆的唯一形式。实际上，我们可以证明这种知识的使用范围是有限的。因为人们所要学习的大多数东西都要求人们对于那些互相联系的论证要点进行回忆。这时，用于保证对词表进行机械记忆的一般记忆术可能就不再起作用了。

假如元记忆的演示研究增多的话，那么这些研究会集中并成功于我们甚至缺乏基本信息的那些领域。例如，测查儿童是否了解自己在书面或口头交流中准确无误地表达出自己思想的能力，区分什么情况下需要重视记忆，什么情况下需要重建记忆，区分什么情况下需要审慎的记忆，什么情况下不需要这样的记忆，以及对记忆过程中其它智力活动进行类似的研究。

我们之所以对儿童或成人的元记忆了解得不够，原因之一是我们对于成人解决这种问题知之甚少。对于回忆和分类组织有了足够的了解，我们才能知道儿童是否也知道这些。对于较复杂的记忆技能知之不多甚至一无所知，则很难知道儿童应该了解什么。科林斯等人(1975)曾高明地考查了能言善辩的研究生进行苏格拉底对话游戏的全部技能。发现，所需要的(技能)和有条理的成人或聪明儿童面对不同的记忆情境时的技能类似(布朗，1977b)。研究各种技能应该有一个良好开端，因为任何学生都要尝试技能，每一个学生都必须了解某种活动及其最终目标的效果和局限性(布朗、斯迈利，1977a)。对于成人的这种理解能力有了较多的了解，我们就能更好地评价不太成熟的学习者还不知道哪些，需要知道哪些以及通过训练可以知道哪些。

C.发展趋势

应该注意儿童关于以下几方面的知识：(1)基本的机械学习策略，(2)记忆和知觉之间的区别，以及(3)不太完善的机械记忆

能力导致了这样一种印象，即认为元记忆的发展是迅速的，而且到了三年级元记忆的功能就完善了。这不是说任何人都认为到了这一时期发展就完成了，但是令人鼓舞的是在这个年龄存在着一致性，这时儿童获得了所研究过的特殊子集技能的知识。有一项使用克罗伊特泽(kreutze)等人问卷的研究 得出 了 令人印象深刻的证据。这一证据表明在三年级前后对于一套题目中的大多数题目的作业的最高水平。而且，大多数实验研究表明，在三年级或三年级以前多数儿童获得了各种觉察。因此，不论回忆的准备状态还是注意广度在二年级以后都没有什么改进(弗雷沃等,1970)。在学习时间分配任务中，在一年级和三年级之间有某种改进，而三年级的行为很象大学生。莫伊纳汉(1973)发现在预报分过类和没分过类的表这种任务的难度时，三年级和五年级之间没有什么差别。实际上，如果在许多现有的元记忆研究中排斥年幼的团体(比如幼儿园或一年级的样本)，那么就得不到可靠的发展差异。这个模式似乎很有跨任务的一致性，到三年级，儿童对表的机械记忆知道了很多。

然而，在任务比较复杂时，比如在判断一段散文的难度或判断课文的不同部分的重要性(布朗和斯迈利,1977a,1977b)时，为了更有实效，应该把年龄更错后一些。统观全文我们自始至终试图说明任务难度作用的重要性。评判儿童知道或不知道：是否具有元认知洞察力依赖于任务的难度以及任务要求与儿童现有认知技能的匹配。一个对组织了解很多的儿童，在组织的基础是分类法时，对于课文材料的组织原则却可能知之不多或根本不知道。

关于发展趋势的另外两点后面将作较为详尽的阐述，在这里仅简单地提一提。第一，克罗伊特泽等人(1975)在提出一种见解时得到了一个重要的结论。他们认为中学生所掌握的关于记忆的某些方面的知识几乎满足不了他们如何完成这种任务的需要。例如，

丹纳(Danner, 1974)指出，关于课文中重要结构特点的知识在很大程度上位于为了补偿的目的而选择适当线索的能力之前。同样，知道某种积极的策略有助于回忆但并不一定意味着儿童自己能够选择、使用这种策略。以后的发展可以以儿童关于记忆的知识以及儿童在记忆方面做了些什么这二者之间不断增长的一致性为标志。在预报和作业那一部分中，我们将回过头来谈这个问题。

第二，关于元记忆发展资料的研究有一个特点，就是强调关于认知的自知的早期信号和元记忆初期预兆的出现。这一特点表现于在尽可能早的年龄阶段寻找元认知的证据，在擅长此道的研究者的巧妙研究中，这种努力得出了结果（韦尔曼、里特、弗雷沃，1975)。由于美国发展心理学"反结构主义"影响的扩大，这种研究得到并将继续得到鼓舞，曾有人寻求某种证据以反对皮亚杰关于学前儿童能够进行理性运算的观点(一种曲解)，反对学习中介理论关于学前儿童可以思维的观点(一种曲解)，结果证明了学前儿童的这种智力本身已经形成了。因此，虽然新皮亚杰学派的许多"游戏"所提供的儿童出现具体运算的年龄阶段越来越早，然而却没怎么强调借助于严格的成熟理解标准来指明早期觉察的界限。在对学生的理解健全性做出真实的评价以前，只能在表面上用他们的反应这种作法作为所学概念智力理解的指标(布朗，1976a)。这种作法也影响了所出现的元记忆文献，因此，在学校面临着"儿童多大才能理解"时，应该更多地强调"上限检验法"。

D.预报和作业

赞成研究元记忆发展的一种最有说服力的论证认为，一个人对于记忆本身了解些什么和他如何进行识记之间一定有紧密的联系。如果能证明儿童不知道背电话号码的用处而不背电话号码也就不足为奇了。同样，如果儿童能意识到运用策略的重要性，那么

就可能会采取富于策略的行为.这样严格一致的证据还很缺乏.很显然，对于记忆和元记忆二者之间关系的系统研究还为数不多。我们所掌握的证据大多数来自事后考试，而这些考试不打算说明预报和作业。到目前还不能足以证明一个人对于记忆了解些什么和他如何进行识记之间的紧密联系。

先来考虑一下我们拥有的资料。萨拉塔斯和弗雷沃，莫伊纳汉检查了与分类和回忆有关的知识。但他们都没有发现直接的联系。莫伊纳汉发现觉察到类别的作用与关于分过类和没分过类的表的实际作业无关。同样，萨拉塔斯和弗雷沃发现，还没掌握分类的一年级学生象已经掌握了分类的学生那样表现出分类有助于回忆。早期的研究都不主张把预报和作业完全分开。萨拉塔斯和弗雷沃发现教儿童记忆可以增加关于元记忆问题正确回答的数目。莫伊纳汉发现，在力求主动记忆之后(而不是之前)对元记忆问题可做出更好的反应。记忆的经验对于元记忆觉察确实有一些作用，但并不是人们所喜欢的直接影响。

预报和作业之间缺乏直接联系这一点在正常的和可教育的迟钝儿童中得到了很好的证明。要求儿童预报模仿活动能产生好的回忆，(灵活策略的优越性)在所有心理年龄6岁和8岁的儿童中都得到了表现(这种现象与不太灵敏的学前儿童和一年级儿童形成对照)。让儿童看了模型以后，立即给他们确切的任务让他们自己完成,只有少数人能自己运用有效的策略(参见表11)。到了三年级，大多数挑选来的儿童都能指出最有效的策略。但是,甚至再大一些的儿童(预报和作业之间的)联系仍不很完善。克罗伊特泽等人的问卷题也表现了这种情况。那些预计学习时间越长，学习越主动将导致更好的回忆的学生不一定说他们能按照自己的方式去做。

弗雷沃和韦尔曼指出，所以没有找出元记忆和记忆间理想的关系，原因是多方面的：

假定一个人认为分类刺激比非分类刺激容易回忆，他就一定是用分类作为贮存的策略并接受明显可分类的刺激吗？完全不是这样。他可能对分类有所了解，而却考虑在当时情境中其它可能更好的方式。他可能在琢磨一个易于使用简单审视作为贮存的表。如果问到分类，他可能有充分的了解，认为分类是一种好的策略，但是并没有充分考虑到自己就应使用这种策略。最后，在元记忆和可归因于原罪的记忆行为之间，毫无疑问地存在着距离。道德行为和道德信念并不一定永远一致。同样，人们并不总是按我们认为最有效的方式检索信息或准备将来的检索(p.p27—28)。

象弗雷沃和韦尔曼一样，我们认为一定“存在着元记忆判断的发展，存在着记忆行为的发展，存在着二者之间的一致性的发展。”但是，对于支持这种说法的证据我们都知道得不多。上面所罗列的全部实际材料，都来自心理年龄为8岁或8岁以下的儿童。在这点上，元记忆和记忆之间的联系是微弱的。在模仿研究中，我们特别注意到在年龄较大的儿童中所期望的一致性的增长。当然，日趋成熟的元记忆的觉察和实际记忆行为的改善这二者之间的发展仍有待于进一步研究。

重复一下弗雷沃和韦尔曼(1977)的另一论点：“因果关系链索可能是更清楚、更独有的元记忆→稍后发展的记忆行为。”但是，这种发展是如何进行的呢？是刚刚萌发的元记忆觉察先提高了记忆行为的效果(元记忆→记忆行为)呢，还是经验的增加和记忆一起导致了元记忆的萌发(记忆行为→元记忆)呢？还是一种复杂的二者相辅相成的过程(记忆行为→元记忆)呢？我们对于认知所了解的一切都会指出是第三者，因为不能指望在反复体验不同的记忆任务之前借助于天赋使儿童具有元记忆的知识从而变得开化。关

于记忆的知识和行动的协调可能是低年级以后发展的精髓。关于这种假设越来越协调的研究几乎还没有开始。

经验的性质能影响这种认知发展无论在理论上，还是在实践上都是重要的。弗雷沃和韦尔曼认为，学校和学校任务的一般经验可以提供这种发展的原动力，但是，缺乏世故的高年级儿童，在评价自己的能力时可能认为这种直接的影响不是很有效的。实际上，可能是这样一种情况：这种智力理解上的一般改进不可能教得十分明确。可是，对于那些对教学心理学感兴趣的人来说，这一领域的大门是开放的，以便人们努力寻求有效改进的基本经验。如果有人对那些如果不进行明确的指导就决不会获得这种悟力的迟钝学生感兴趣的话，那么这一点则是特别重要的。

E.觉察的测量和标准

由于元认知研究还不很成熟，所以还没有制订出用于评价元记忆觉察的精确的量度这是不足为怪的，但是，其它有关领域中令人痛心的事例很多，我们应从别人的错误中吸取教训，这才是比较明智的。我们从有关的认知发展领域中了解到，把问题和儿童的言语反应结合起来作为儿童知道了什么的指标。儿童说他做了什么，想做什么不一定和他做什么有关；所以，依赖言语反应并以为正确，实际上是一种冒险，虽然儿童赞成这样做。正因为在元记忆领域中大部分硬(软)材料是这种言语的自我报告，所以评价这些资料的标准问题是一个很棘手的问题。

测量元记忆判断的问题是很敏感的，因为我们关心的不是儿童正在做什么，而是儿童在这样做时想的是什么和为什么这样想。了解儿童知道什么的直接方法是向他询问。这种方法的某些问题事例可证明是说明性的。有一位从事模拟研究的实验者在她7岁的儿子看完录像带后问他是怎样学习画录像上的画的。他毫不犹豫

地答道要盯着看，如果必须记住的话就总盯着看。有一系列图画，他仔细地把所有图画分类，把不同的类别分开，然后开始有系统地细看这些图画。问他要记住什么，他说，他正在注意看这些图画，正如他自己说的那样注意看图画。

只考虑口头回答的表面价值而陷入窘境的轶闻在文献中并不多见。或许最富于说明性的是布朗等人在觉察指数是儿童达到其能力的第一个指标时或在允许儿童继续估计表直到最大表长为10项时所发现的儿童注意广度上的差异。如果在第一个答案之后就停止并对他们的答案作出恰如其分的评判，那么许多儿童都会说一连串5道题难做，而却说一连串6道题并不难；或说一连串7道题太多了，而说一连串8道题还可以。不论儿童究竟回答的是什么，这一点可以告诉我们关于儿童记忆的某些特点，也确实提醒我们，测量觉察时不能轻易地只轻信一个口头答案。

克罗伊特泽等人的问卷要求多重回答，其中包括适当的答辩，试图用这样的方式来解决这一问题。表明觉察和适当答辩的一个“是/否”答案之间的匹配显著地随着年龄而增长。要求适当的答辩能解决这一问题吗？库恩(kuhn，1974)从皮亚杰守恒研究的角度考虑了这个问题。很显然，不只有一个思想流派。布雷纳德(B-rainerd，1973)认为用答辩来评价儿童对一个问题的理解是不合适的，因为我们假设操作性先于用语言来表达这种知识的能力。第Ⅱ类错误的危险也成为一个问题，因为许多儿童可能很好地掌握了必要的认知技巧，但是却不能恰当地把它们表现出来。布雷纳德主张使用“是/否”，“相同/不同”的回答方式，但是，库恩指出，任何对偶选择法对发展上具有敏感效果的反应都很敏感，我们知道，反应倾向效果是越来越微妙的。进退维谷的原因在于，对反应要求答辩则导致了第Ⅱ类错误的可能，而依赖对偶反应则有犯第Ⅰ类错误的危险。库恩的解决方法是使用集中操作。应诱导被

试做出尽可多的不同反应，根据所有图画提供的(材料)来评价觉察的程度。

由于大多数元记忆资料中反应也是对偶抉择或答辩，所以在这个领域中似乎值得研究这同样的方法。应设法取得关于觉察的多种不同测量，如果做不到的话，至少也应避免把某一种反应作为儿童的知识的唯一测量。很显然，还要考虑所获得的测量的信度问题。

在策略—选择模拟任务中，有23%的正常 4 岁儿童，44%的心理年龄 6 岁的智力迟钝儿童和28%的心理年龄 8 岁的智力迟钝儿童当使用多方测量时对所喜欢的策略的选择不一致。这就说明我们有必要得出信度资料。值得注意的是，布朗和劳顿发现，在某种情况下对自己了解的感觉很敏感的儿童对该任务的变式并不一定表现出同样的敏感性。尽管在那些研究中有些严肃选择的问题。如果还需要做进一步的告诫，可以根据你自己的观点来看某些儿童的“不诚实”或“创造性的灵活”。这些儿童在回忆测验中故意不好好做，以便可以玩流行的“知道感”打赌游戏。当然，他们对于那些能很好地回忆起来的题目的认知准确性有着可靠的知道感。所以，我们同意贝尔蒙特和巴特菲尔德的观点，即在这个领域中，测量是一个至关重要的问题，而且，在寻求儿童对自己心理运算觉察程度的定量和定性上，我们赞同会聚运算(参见第Ⅴ)。

F.训练和迁移

目前对元记忆感兴趣的另一个动机(当 然 也是我们对这一领域感兴趣的原因）起源于这样一个争论，即在那些不按这种方式思考的人中试图训练特定记忆策略的成功限度问题。这个争论很简单：如果幼儿对于记忆术的作用一无所知，他们为什么能从教学中受益呢？如果训练复述，他们就复述，特别是如果情况不变，

他们能得到连续的提示的话。但是，为什么如果在活动的原因还不清楚时要希望儿童很有见识地运用自己的新的技能呢？

这些把我们引向了这样一个有趣的问题：我们应训练作业的哪些方面，是周密细致的记忆技能？还是这些技能的执行控制？巴特菲尔德和贝尔蒙特提出了这个问题，比较了他们认为布朗，巴特菲尔德和万博尔德、贝尔蒙特所持的不同观点，他们把我们的观点视为根据特定任务的答案来抉择个别策略的训练，期待着执行功能将做为掌握适当技能子集的结果面出现。反过来，他们感兴趣的是执行功能本身。虽然这反应了不同的主张，但是，当我们考虑到促进训练大纲必需的实施步骤时，这个不同的重要性就失去意义了。因为很难想象如果没有一套控制策略的话如何训练策略的执行控制。可以建议儿童“事先计划”，“使认知活动实用些”，“使用唯一正确的计划，而且一旦不需要了，就停止使用”等等，以便教育儿童凡事要有计划，这虽然是可行的，然而实际上，这似乎表明在试图引导监控这些策略行为之前必须先就得训练某些记忆技能。巴特菲尔德和贝尔蒙特基本上同意这种观点，因此，意见上的不一致实际上是人为的。他们指出：“由于控制过程(技能等)是执行功能的议题，所以，控制过程应该视为行使功能的最有希望的指标。因此，控制过程的基本测量应先于其监督的测量。”我们认为他们对我们的观点的描述欠妥，因为他们在1974年的文章中指出：“一旦教会了合用的技能或技能子集，下一步就要提出训练落后儿童的方法，使他们学会监控自己的策略运用，并实实在在地评价任务要求、他们自己的能力和特殊技能指令系统之间的相互作用。”

仔细分析以后就会发现明显的意见分歧：尽管这两种观点都主张把预备技能的训练作为研究执行控制的基本必要条件。然而，在侧重点上所表现出来的不同似乎能反映出训练方向上的更为本

质的差异。贝尔蒙特和巴特菲尔德在后来的文章中似乎改变了他们的侧重点，要求对训练情境和所有附随的转化测验都进行细致的任务分析（例如训练特殊技能）。我们现在也转移了我们的重点，强调训练解决问题的一般途径，而不是强调训练个别的策略。

这种争论需要仔细推敲。贝尔蒙特和巴特菲尔德极力主张对逐个实验室任务以及被试可能使用的策略进行细致的任务分析。他们为改进累积复述、改进快速完成策略（这种策略被认为是他们的范例中的最佳策略）而进行了十年的努力，这很好地说明了所能进行的详细分析的程度。这种详细的任务分析的作用是很清楚的。对于这种最佳策略的各个方面都了解透彻，就能把儿童训练得更有能力，就能诊断出为什么训练未能达到最佳效果，还能把作业水平至少提高到未受过训练的成人水平。

这一系列令人赞叹的研究，特别是巴特菲尔德等所报告的研究，说明了用来训练记忆术策略的详细的任务分析方法的作用。如果训练的目的是要看一看如何使儿童的行为接近成熟的作业，那么这种方法是很有成效的。从理论上来讲，这些资料是无价的。因为这些资料证明了关于"结构限度"观点的解释是不正确的。如果训练失败了，我们不应该认为这种失败是由儿童的某些基本能力的局限性所造成的，因此还应力求改进训练。如果所希望的最终结果是改进训练任务本身的作业的话，那么任务分析方法从实践的观点来看也是非常宝贵的。戈尔德(gold)对严重落后的被试所进行的研究是一个杰出的例子。如果通过详细的任务分析把一项智力任务分解成易操作的小单元，那么严重智力落后而被送进专门机构的人也能够很快地被训练得能完成复杂的"装配"工作。这种训练过程的目的是要使训练任务迅速无误地完成，这样，就可以使至今还没有工作机会的人运用这些技能自食其力。

从事认知教学的目的一般地来说是不一样的。不能把目的看

作是某个特殊任务的优秀成绩，倒是所期望的最终结果影响着理解的一般提高，影响着进一步详细说明的需要，这种目的可以进一步得到理论上和实践上的维护。从理论上来说，如果没有广泛地迁移的证据的话，训练的结果可能仅仅导致机械记忆规则的掌握，而没有产生真正的"结构变化"，也没有使儿童对于客观世界的了解有全面的增加。因此，至少从两个方面来考虑"反结构主义方法"。在某一任务上表现出的类成人作业，对那些热衷于证明智力的不成熟不一定影响某一任务完成的人来说是一个充分的证据。极端的观点是主张只要在适当的训练条件下就首先教给每个人所有的东西。另一种结构局限观点和皮亚杰的观点相似，换句话说，年轻的思维者的推理能力有一定的局限性，只有达到了适当的成熟水平，某种特定任务训练才能影响他们的能力。我们可以设法促进智力的成熟，但是训练能起的作用并不太大。在记忆训练领域中，赞成"结构主义"这种更为保守形式的人在寻求概括化，以作为训练成功的指标。

这一问题并没有为贝尔蒙特和巴特菲尔德所忽视，但是他们心目中似乎有不同的最终结果，他们的目标是"在每一个可测量的方面都把儿童的某种特殊任务的作业提高到成人的水平"。这一目标是借助于一个有效的计划来实现的。关于这个计划，他们说到"任何人只要按着这个计划把思想集中起来就能很好地回忆"。有了这个计划，幼儿就能执行得"十分准确，就好象是他们自己发明的一样"，"教学研究者及其年幼被试显然从事着同样的事业。而且，谁做出执行决定是无关紧要的，只要这些决定做得好"。最后，他们断言，教学法的惊人之处在于教给儿童"怎样思考和思考什么"。

"怎样思考和思考什么"这种说法囊括了这两种方法之间的差别，因为在训练儿童如何思维和训练儿童思维什么之间存在着很

大的鸿沟。怎样认识和认识什么的争论在教育哲学的历史中起了很显著的作用，这个争论至今仍很活跃。这一问题与格里诺的教学的认知目标和行为目标之间的区别相同。训练一套规则或秘诀对改进任务的完成起很大的作用，但是，我们认为，如果不同时去了解使用这些技能的理由，那么，其结果未必能改善儿童怎样思维的知识。我们在这里无需详细介绍发现学习与机械学习之间的长期争论，请读者参见第Ⅲ部分D·3的例子。

虽然大家对把迁移作为认知活动全面改进的标准已深感兴趣。但是，这篇文献指出了这一领域实验研究的不足。之所以产生经验上的不足，不是由于缺乏兴趣，而是因为要进行这样的研究需要花费惊人的时间和精力。贝尔蒙特和巴特菲尔德主张把他们的任务分析法再延伸一下，以便用来研究迁移："准备研究迁移的人对在训练中所使用的任务和为检验迁移而使用的任务都必须有透彻的理解。"让我们再来强调一下他们所说的"透彻"，因为他们主张训练任务和迁移任务都必须接受他们在累积复述任务中大量使用的那种深度分析。如果不能对迁移任务进行真正的深度分析就会使研究者没有办法解释为什么迁移不能成功。不能适当地迁移，可能归因于受训者不能看出所要训练的行为和新任务之间的关系。这是颇为常见的解释。也可能是由于受训者不能执行无论是受训者还是研究人员都不能完全理解的迁移任务中的某些其它成分。

那么唯一的答案是去考虑从一个经过充分分析的任务到另一个任务的迁移。就我们所知，仅有一项尝试接近了这样的标准，但是没有成功。在这项尝试中，训练落后儿童不断地练习保持记忆任务，并且在六个月期间不间断这种活动。然后把贝尔蒙特和巴特菲尔德的探查回忆复述任务给这些受过训练的落后儿童作为迁移测验。结果，首次参加实验的儿童和受过训练的儿童的分数

相同，而且都很低。更重要的是，在这两组被试中均没发现复述活动(停顿方式、观察到的外在行为)的证据。实际上，可以认为迁移情境不太理想：复述训练后过了一段时间才进行迁移测验；以及保持记忆表由4个题目组成而探查回忆测验却由6个题组成等，是实验不理想的原因。但是两种任务都要求连续背诵三个一组的项目，都要求训练策略，但是甚至在迁移的前三个项目中也没表现出活动。我们不能就以此作为有力的很据，而认为这样的迁移不能实现，然而实验的结果确实不怎么理想。我们对贝尔蒙特和巴特菲尔德的观点的解释是，在获得可靠的结论以前需要进行第一个任务的详细任务分析，第二个任务的详细任务分析，详细的任务分析后进行两个任务中的非共同成分的训练。迁移条件的详细分析以及……。实际上，我们必定遇到时间问题，应提出训练者和受训者有效认知的原则。如果训练的目的是影响概括化，必须考虑到任务分析法在实践上的局限性。

正如我们所看到的那样，对所有可能的迁移任务都进行详细的任务分析实际上是不可行的，所以我们必须提出一些别的方法，或者至少要有某种方法胜过或者减轻这种费时的尝试。我们认为合适的迁移测验的标准是：那些在前测中能自发地采用训练策略的人在各种迁移任务中也能尝试着使用这些训练的策略。我们可以从几种途径获得这些信息。例如，使用累积复述。可以问一问那些自发地采用某种策略成功地进行复述的人，在什么地方，为什么他们想到运用这些策略。进行一系列这样的调查，就能发现典型的复述情境，那么，差不多任何人遇到这样的情境就都能进行复述；在一些近似的情境中对于复述的合适性有不同的意见；在一些相差较远的情境中，对于复述结构有所了解的人没有一个能使用这种结构的。获得相同信息的第二种途径是设计一套任务(典型事例，接近或不接近的情境)观察自发地执行任务的人在什

么时候、什么情况下使用这种策略。最后，在任何研究中，任何获取理想信息的方法都要考虑那些前测完成得很好的被试在后测中的概括作业。

布朗、坎皮尼和墨菲1977年进行的广度估计的研究应用了第三种方法(参见第Ⅲ部分E.2)。最初，现实的儿童能很好地完成所有的概括测验。这表明了他们对于测验任务的要求和自己的能力都很清楚。然而，经过训练以后，这些儿童却不能很好地完成这些概括测验，这就表明了训练的局限性(参见表16和表17)。虽然这可能表明前测完成得很好的儿童与其他儿童相比，在某些内在的认知因素(这些因素与能否很好地完成前测和一般作业有关)上存在着差异，但是这也没什么关系。如果训练的目的在于改进作业，以便赶上那些能很好地完成训练任务和一般理解的儿童，那么一开始就能完成得很好的那些被试所达到的标准显然就是成功的训练应达到的标准。

从我们的观点来看，训练的目的并不是为了使儿童更象成人那样完成任务，而是为了使他们在某些类似的情境中象成人那样思考。如果这是较为理想的目标的话，为什么不训练①概括化，或②可能适合于各种情境的技能呢？就我们所知，没有哪些研究不设法进行概括化的训练，实验者甚至用概括化来作研究的标题。但给我们的印象是，参与这类实验的儿童被当作“敌人”来对待，而没有被当作教学方法的合作者。我们愿意指出即使儿童能象成熟的学者那样自发地进行概括，然而，为了帮助缺乏能力的儿童，下一步要暗示他们应该或更好是尝试直接的概括化教学，否则，不成熟的受训者就不能进行概括。

这样的训练必须和特殊的规则学习相结合，不然的话，我们只能围绕着巴特菲尔德、贝尔蒙特和布朗的争论兜圈子。一旦我们依照前两个标准——这种策略的使用和维持，训练掌握了记忆

技能，难道就不可能插进某些特殊的概括训练了吗？例如，可以告诉儿童，所训练的行为在许多类似的任务上都能对他们有帮助，窍门就是要知道是哪一类任务，然后，让儿童完成各种原型任务，在这种示范情境中使用这种策略。这时，应考虑下一步的训练任务。而且，应该讨论和证实训练行为不合适的原因。最后，可以让儿童进行概括测验，这一测验包括新的原型和下一步任务。通过这样的测验，可以考查儿童是明智地运用策略，还是不明智地运用策略。这样的方法是否能成功，我们不太清楚。但是，学习缓慢的儿童需要明确的训练，对于这一点，我们获得了令人难忘的证据。这当然是值得试一试的。

关于迁移的最后一点，重新考虑我们试图训练的技能的类型可能是明智的。成熟的记忆者练习得有多么勤呢？如果儿童因看不到训练策略的作用而不能概括出这种训练策略，那么这就可能是关于这一努力的一个真实评价。终究，他们都告诉我们他们写下了电话号码。

如果有一种替代策略的话，那就是去训练能对问题解决情境具有广泛概括性的原认知技能。有一点应记住的是，我们选的文章都有"检查"、"计划"、"质疑"、"自我测查"、"监控"等标题。这些技能是跨情境的。也许我们能训练儿童在解决某一个问题之前先停下来思考一下；向自己或别人提一些问题以断定自己对所要解决的问题是否有所了解；不是通过问"是否正确"而是问"是否合理"这样的问题来检验自己的答案；监控自己的企图，以便学会观察这些企图是否有效，是否值得试一试。我们知道有许多问题与这些建议有关。然而，在完全缺乏资料的情况下，我们无从知道这种训练的干涉与是否能多少改进发展上比较幼稚的儿童的记忆知识。但是，考虑到以往诱导概括化令人沮丧的失败，为实践这些建议花些时间和精力还是值得的。

G.训练的局限性

关于训练研究的总的观点是，我们必须说明因年幼、缺乏经验或智力低下而造成的训练结果的局限性问题。智力及其如何定义的问题是很困难的问题，比较研究的某些方法论及哲学问题我们在下一部分再进行探讨。为了进行如下的讨论，明显需要说明的问题是：不论何故，个体对教学要求和测验情境的适应性是不完全一样的。可教育的智力迟钝儿童正是由于缺乏这种适应性才被挑选出来，接受学校的特殊教育和实验室的加强训练。为了阐明我们的观点，如果给智力下一个工作定义可能是有益的，我和我的同事都使用这样的定义。我们认为目前发展上所出现的关于"能力"或"结构的"局限性的争论在很大程度上是因对所用术语缺乏明确的定义而造成的。关于能力的发展变化的讨论之所以常常是没有道理的或混乱的，最主要的原因可能就是由于所使用的关键术语是随便的，可以换来换去，而且常常是用词不当。对于这种讨论所使用的比喻手法必须了解得一清二楚才能理解这样的讨论，这些比喻多半都意思不甚明了，而且有的作者所使用的比喻还前后不一致。这里，我们不介入能力/非能力局限性的讨论，关于这方面的情况，读者可参考奇和赫坦洛克(Huttenlochen)以及伯克(B–uke)1976年的论文。

为了表明我们观点，我们倾向根据执行功能给智力下一个定义。为了扼要重述，考虑到属于现代记忆理论中的执行功能的一系列责任。有效的执行能承担常规选择、应用和控制的复杂协调。其责任是监控，检查以及评价根据一些有效的标准所选择的常规，并推断出目前的知识状态以及该知识状态和所要求的目标之间的匹配。执行功能还应对使用可用于该系统的方法达到目标的概率做出估计。执行功能的特征就是我们主张作为训练项目的特点：

检查、计划、监控等等。然而在这里实际上意味着我们希望训练有效思维。有效地进行思维是很好的智力定义。

关于这一点，要考虑一下当机器进行思维时，如何给智力下定义。那些想给发展不成熟儿童或机器的智力运算下定义的人所遇到的问题的相似性是惊人的。穆尔(Moore)和纽厄尔借助于两个标准给机器智力的本质下定义，第一"如果在合适的时候S使用K；那么S就理解了K"，这是知识和对知识的理解二者间的区别。不成熟的思维者达不到我们的实验室任务的这个标准，因为这是个坚持性概括化的问题。

适当地使用程序的能力是机器智力的本质，这也是人类智力的合理定义。人们正是用是否能合理地使用某种技能来定义智力行为，而不是用训练情境中那种技能的适当作用来定义智力行为。必须自我生成运用活动的执行决定。假如由外界代理人来选择常规的话，那么就是这个外界代理人(教师，实验者)而不是儿童是智力的行动者。因此，我们认为由谁来充当执行者是很有关系的。假如目的在于改进某一特殊任务的作业，那么由谁来做执行决定是无关紧要的，然而，如果目的在于改进思维，那么由谁来做执行决定则是至关重要的了；S必须凭自己的意志适当地使用K。

穆尔和纽厄尔进一步考虑到把适当使用知识的能力的范围作为对机器智力的评价："理解无论在范围上(使用知识的适当情境的种类)还是在直接性上(达到理解前所用的时间上)都是不完全的。"这些标准类似于雷斯尼克和格拉塞把人类的智力定义为一个人学习周围环境中重要事物的速度和效率。因此，可以用这种观点来考虑训练的效率。儿童对训练的反应速度和效率究竟如何？在以范围(广泛概括)和直接性(没有额外的提示和训练)来测量效率时，他们迁移信息的效率如何？我们愿意指出，如果达到概括化的合理尝试没能成功的话，甚至当迁移是合适的(标准的)和对

于概括化有明确的指导也不能成功时，就证明存在着发展的局限性，我们可以把这样的局限性看作是智力局限性的合理解释。

雷斯尼克和格拉塞还认为，智力是在没有直接的或完整的教学时的学习能力。因此，如果概括化只有靠直接的概括训练才能获得的话，我们就没有必要把这种成功看成是对智力局限性观点的一个反驳。因为，如果它只能通过训练才能获得的话，那么这样的概括化就不能达到"没有完整的教学时"这一标准。因此，没有理由认为如果在某个领域中必须有详细的教学才能获得迁移的话，那么迁移在另外的领域会自然地发生。

因为我们举了有利于"反结构主义"观点的例子，所以我们介绍了这些论点，也就是说，由于早先的研究强调策略的训练，才假定我们认为发育中的儿童和较成熟的思维者之间不存在"能力"方面的差别；他们在作业上所反映出来的差别是由于不适当地运用(或根本没有运用)记忆策略造成的，而这种策略的运用是可以训练的。这种论点依赖于能力一词的定义。我们愿意再一次表明这样的观点，当一任务的专门训练没获成功时，求助一个没有详细说明的"能力局限"在逻辑上很难进行辩护，而且在可采取的训练方法得到验证之前，这样做尚为时过早。现在还没有有力的证据表明可根据系统结构(如短时记忆)方面，结构单元中的空间数量(在短时记忆中存贮槽的数量)方面或者这些系统中信息持久性方面的能力差异从而把不成熟的思维者与成人思维者区分开来。

然而，这不是说，发育中的儿童不受他们的中央加工能力局限性的束缚，各种局限性不能归因于简单的能力概念。薄弱的知识库(长时记忆)的影响就完全能证明许多已报告过的发展上的困难。长时记忆是规则，策略以及运算的贮藏所，而规则，策略和运算又可以为有效地运用有限的能力系统创造条件。另外，儿童的知识库至少在以下三个方面是不足的：①知识库所包含的信息量；

②这些信息的组织和内在一致性；③形成知识库可用的线路数目。这些差异使儿童的信息加工能力在几个方面表现出局限性，甚至象短时记忆中的阅读信息这种简单的情况也表现出局限性。象检索的流畅、编码的速度、命名和再认这些基本的认知过程无一不受到一个薄弱的知识库所表现出的上述局限性的影响。

知识库的局限性显然不会因短时的训练和多么巧妙的教育干涉而改变。人们不能指望在3岁，或者甚至22岁时用些轻松的课程来丰富知识库。当然，这种“结构”的局限性不是固定的，随着年龄的增长和经验的丰富将会导致知识库的重组和丰富。教育干涉的形式可以是把知识贫乏、经验不足的儿童置于一个具有丰富的刺激和经验的环境中，这是许多幼儿早期教育计划的一个基本指导思想，但是不想出一套达到教育目标的详细训练计划是很困难的。鉴于目前我们对知识库的了解有限，这就影响了训练。

另外，局限性还表现在对于智力低下儿童训练的结果上。我们认为完成适当作业所需要的训练范围和用迁移的范围和直接性所表示的训练效率都是智力的指标，尽管在得出这一结论之前无论是任务还是训练的文化相对性都需要进行仔细的研究和调查。

智力的另一个特点是，不仅已学会的技能能适当地用于新的地方，而且还能进行创造性的改造以满足新任务的要求。此外，仅根据已有的知识就能创造出新的情境。雷斯尼克和格拉塞最近对智力和发明才能问题进行了讨论，在此我们不想对其细节多加阐述，只想指出关于智力思维的标准是我们要求已掌握的技能向原型迁移任务自发概括这一点证明得很不充分。

前面的所有论述都是这样琐细。首先，显然正在发育中的儿童与成熟的问题解决者之间有着功能上的差异，因为，如果不是如此，那么教育的干涉就不必要了。第二，并且是更重要的一点，人们认为关于这种发展差异的原因、类型，和对矫正教学的敏感性

对实际的训练问题没有什么影响。意见的分歧是理论性的而不是实际性的，是描述性的而不是规定性的。我们象任何人一样坚信为了提高问题解决能力，特别是迟钝学习者的问题解决能力，应该不遗余力地进行教育的干涉。教学心理学家的一个主要责任是设计出越来越巧妙的训练计划和措施，以便提高作业水平。另外，我们对于设计完好的训练的成功和值得花时间和精力改进作业、获得成功的现实意义始终抱着乐观主义的态度。我们做了这一部分的论述是因为我们认为训练的结果是有限度的，所以，我们愿意与极端的反结构主义的观点划清界限。

V. 记忆、智力与教学

在记忆发展的领域里，我们把研究的重点放在可训练的，尤其是那些学习迟钝的儿童身上。这方面的研究是在实验室环境中进行的，其目的是检验具有理论意义的假设。尽管这些研究还是比较初步的，然而具有实际意义的信息却是垂手可得的，这是一条基本的原理。我们知道，要设计出详细的课程，就需要经过一定的职业训练和技能训练。尽管我们还没受过这些训练，但我们仍然期望我们的一些训练技术能够在教学环境中得以实施。在这一节，我们首先讨论常规教学、学校以及它们对于元认知发展的影响。然后再探讨一下关于学校中的条件不利儿童的问题。最后我们将提出能够被证明是行之有效的教学方式。

A. 正规学校教育的作用：各种文化间的比较

从儿童的自我意识在认识过程中作为一个积极的因素第一次出现开始，直到建立起复杂的记忆执行机能（高中生或大学生有时才表现出这些机能），人的记忆的发展经历了一个重要的时期，

人的记忆技能上的主要变化也都发生在这一时期。而这一时期又恰好与大多数西方社会的正规学校教育时期相吻合。所以，对这一发展时期进行研究是很有意义的。上述这两个时期的吻合是否意味着正规学校教育与人的记忆发展之间存在着某种内在的联系，或是表明教育的作用大于成熟的结果呢？解决这一问题的唯一有意义的方法就是以某些特定的文化背景作为参照，在这种作为参照的文化背景中，正规学校教育程度与实足年龄两者的作用不象在欧美那样无法区分。通过对跨文化研究文献的分析，我们可以找到许多本章已讨论过的技能与文化因素的关系。通过这一分析，还使我们注意到了一些哲学上和方法论上的问题。心理学家在研究工作中，要根据所要获得的资料的性质和特点等因素来确定所要进行的实验的类型，而在这一过程中必然要牵涉到一些哲学及方法论上的问题。最后，通过这一分析，还使我们在对所获得的资料作解释时特别注意效度问题。

我们首先来看一些关于记忆效率的跨文化的心理学依据。受过学校教育的人与未受过学校教育的人之间所存在的差别依然是在有意识记时选择记忆类型能力上的差别。这一点，我在本章的前面已经提及。在机械记忆时，一般地应对材料进行组织、复述或分类。面儿童能够自觉地做到这一点以前，对他们进行几年的学校教育是很有必要的。未受过学校教育的人除了缺乏那些我们认为一般人都应具有的常规记忆技能外，在根据不同的实验室任务很快地变换解决问题方法的能力上他们也与受过学校教育的人有所不同。斯克利勃纳和科尔(Scribne & Cole)指出，未受过学校教育的人的一个认知特点是：他们倾向于把通常实验室的学习和记忆任务都当作孤立的问题，每一个问题对他们来说似乎都是新的问题。简言之，人们可以发现他们明显地缺乏学会学习的能力。然而，对于受过学校教育的人来说，他们似乎都有明显的倾

向，即把实验室的问题作为他们平时在课堂上所碰到的问题的范例。将一般的操作及规则运用到与实验室任务类似的领域，这似乎是正规学校教育的结果。

在实验室任务的完成过程中，如果需要用到某种元语言学能力的话，那么我们还可以看到，这两种人之间也存在着类似的差别。例如，儿童在面对与现实生活相独立的，孤立化了的语言时会感到茫然。奥歇逊(osherson)和马克曼曾问过幼儿们一些这样的问题："外面是正在下雨呢还是没有下雨？"或"我手里藏的东西是蓝色的呢还是别的颜色的？"孩子们为了确定这些问题在现实环境中的实际意义而盲目地去寻找经验的支持。他们认识不到一些简单矛盾中的非经验的性质。或者，至少他们不愿在缺乏经验依据的情况下对这些问话作出评价。评价语言需要经验支持的另一个类似的例子，来自斯克利勃纳在受过学校教育与未受过学校教育的克派尔村民中所进行的关于理解经典的三段论法的研究。在这项研究中，她所提出的问题是："所有的克派尔人都是种稻子的农民，史密斯先生不是种稻子的农民，那么史密斯先生是克派尔人吗？"未受过学校教育的克派尔村民拒绝回答这一问题，因为他们在实际生活中没有碰到过史密斯先生这样一个人。他们似乎不能理解这种只由两个陈述命题的结构关系决定其逻辑含义而又与他们的实际生活不同的任务。尽管我们的资料还很有限，但这里我们再一次看到：存在于传统的经院问题解决情境中的某些逻辑思维形式远不是成熟的自然结果，它必定要经过正规的学校教育才能培养出来。

皮亚杰所进行的跨文化实验研究也支持了这一观点，即正规学校教育的程度是决定能否向抽象思维的高级水平发展的一个重要因素。劳埃德(Lloyd)借助另一些心理学依据，认为皮亚杰学派所定义的形式运算的出现，是紧密地依赖于西方式的教育的。而

且,"既然形式思维,命题演算和数学四群皆为西方式思维的产物,那么这一点也就没有什么值得惊奇的了。当然,它们作为成熟的认知发展的目的,其普遍性仍然是一个悬而未决的问题。"皮亚杰是赞成这一提法的。例如,他自己最近提出,在某种特殊的文化背景下,形式命题思维就可能根本不出现。

最后,奥尔逊(Olson)及布鲁纳提出,从口头文化到书面文化的过渡影响着人们认识活动的基本发展过程。在这方面,露利亚(Luria)举了一个很有说服力的例子。这一例子表明,人们在文化方面所培养起来的认知活动类型是由其社会经济环境决定的。这一例子还表明,人的语音书写系统的发明对于人类思维的性质也有着深远的历史影响。在这里我们不能对此问题进行详细的讨论,我们将集中讨论由西方式教育养育出来的三个相互联系的文化传统的重点,即语言的孤立化和形式化问题、强调语言的逻辑功能而不是修辞功能以及强调一般的无上下文约束的规则而不是特殊的经验。

简而言之,一种口语传统是紧紧地依赖于某种受到听觉记忆限度所制约的思维形成的。这一系统很适于处理谚语、谜语、格言等。然而,哈夫洛克(Havelock)指出:"无论是原则、还是规律或公式都经不起可用口头记忆的句法的检验的。然而,活动着的人及所发生的事件等是经得起检验的。口头记忆的诗句(包括叙事诗)都是表达一些应景性的事件。它所涉及到的是一种偶发事件的现象而不是一个诸多原则的排列表。"

口头记忆常指向于韵律、谜语、谚语、比喻以及妙语等,而书面信息则仅适用于对某一事实的意义和影响的分析。当孩子们学习写作时,就应该引导他们克服对具体环境的依赖,即克服为使文章具有某种意义而去寻找经验支持的现象。他们应该学会写那种读者仅靠这篇文章本身就能懂的东西,因为读者在读到他们所写

的东西时，无论在时间上还是在空间上都与作者有一定的间隔，并且读者不理解时也无法向作者提问。应该引导孩子们理解并写出这种明了的，不受语言环境束缚[①]的书面语言。我们知道，在文字材料上，除了纯粹的文学作品外，还有对事物的一般解说及科技文章等形式。而这种文字材料形式不仅依赖于不受语言环境束缚的、意义明了的语言，而且还要求对术语的规范定义及对某些命题的分析。

而奥尔逊认为，文化传统使得语言专门化了，即使语言增强了它的逻辑功能，但同时却失去了它的社会功能和修辞功能。语言的修辞功能包括来自社会的、官方的以及与经验相联系的情境等方面的交流。处于某种文化背景下的儿童必须学会从文章所包含的官方主张和社会功能中区分出它的逻辑成分来。

布鲁纳和奥尔逊认为在文化社会中，对于语言的形式功能及逻辑功能的强调影响了人们的认识方式，并使得对智力所作的定义具有某些偏见。他们还认为，日益增加的跨文化研究文献支持了关于差异假设。某种文化传统中的正规学校教育影响着人们认知发展的过程；那些相对来说未受过学校教育的人的认知活动的类型不同于正规学校教育培养出来的人的认知活动类型。这是否意味着我们可以接受一种缺陷假设，即认为未受过学校教育的人要比受过学校教育的人智力差一些呢？当然不能。因为在我们所处的社会中，探讨这种差异的各项工作无不与智力有关。

对于现有资料作出的这种解释是不合理的。既然思维类型及认知方式的形成是受文化因素影响的；那么我们就不能把与智力

---

① 严格地说，任何信息，哪怕是一个字，都很难真正摆脱经验的影响，因为读者很容易根据常识去消除文章中的模糊之处，或用过去的事例来帮助理解。采用"孤立化"一词是为了在这里帮助作解释之用。此词用的是奥尔逊所确定的含义，即指将书面信息从具体的社会环境中解脱出来。

有关的文化因素排除在外而去想当然地确定智力的定义。假设我们接受了这样一个关于智力的定义，即“智力就是对现实生活环境的适应能力”，那么我们就应当把一些完全不同的行为也认为是对不同文化的适应。所以，认为未受过学校教育的人较笨，这是不符合逻辑的。因为对于某种实验室任务或智力测验，他们是以不同于西方儿童的方式来完成的。如果我们不考虑那些成长中的儿童所必须适应的文化因素，而去孤立地讨论智力，这将是不正确的。

科尔和斯克利勃纳提供了大量的证据来证明跨文化研究中的一些发现并不意味着谁优谁劣，而只意味着差别。他们并且还证明，不论是在传统的社会，还是在教育化了的社会，人们都具有逻辑思维、信息的有效交流、联想记忆以及概括等思维特点。只是根据上述认知活动所采取的形式，而不是这些认知活动的存在与否来区分受过学校教育的人与未受过学校教育的人的。在斯克利勃纳所进行的研究中，未受过学校教育的克派尔人对于完全不熟悉的三段论法问题作了“错误的”回答。然而在这些错误的答案中也包含着许多漂亮的推理过程。这些推理过程是根据题目所提供的依据进行的，是符合逻辑的。我们来分析一下克派尔人认知活动的这一优秀事例。当请克派尔村民以自己的观念把一系列题目分组时，他们就把这些题目根据它们的功能关系进行分组，然后，这些克派尔人被问道，如果一个笨蛋来对这些题目进行分组的话，那么情况又如何？这时，他们把这些题目按分类学的标准进行了分组。十分有趣的是，功能分组法似乎是被普遍地接受的，而按分类学的标准进行的分组法却受到接受正规学校教育新近性程度的影响。在我们的社会中，只有这种按分类学标准进行的分组法才被当作更高一级的智力的衡量标准。这反映了我们将一种活动看作比另一活动更能表现出“智力”的文化偏见。在一份关

于认知发展的跨文化研究的报告中，劳埃德得出了这一结论："人们都具备基本上相似的智力技能，但是这些智力技能都是在不同的文化背景下，以各异的方式形成的，……(并且)没有哪一种背景可以认为比别的背景更优越。"

米歇尔·科尔和他的助手们指出：长期以来，人类学家们所强调的是处于不同文化背景下的各种人之间智力上的共同点，而不是强调其差异；正是心理学家们在强调这种差异。人类学家们的工作主要是对于自然出现的行为所作的人种前的描绘，这与心理学家们所采用的测验和实验不同，人类学家反对心理学家们所采用的传统的实验室任务，不仅借口在西方社会条件下发展起来的实验材料、任务及程序带有种族中心主义的色彩和文化偏见，而且还因为实验本身作为诱发认知依据的背景，对于所归宿的文化背景没有任何生态学意义。

劳埃德、科尔及斯克利勃纳都一致相信，只要对所得结果的解释给以适当的注意，那么传统的实验室任务仍可用于跨文化研究。劳埃德认为："在对某一特殊任务的反应上所表现出的文化差异应该成为跨文化研究的起点而不是目的，并且应该仔细检查关于某一任务的作业，以便确定它是一个有意义的反应还是企图满足一个外邦研究者的随意怀念。"

科尔和斯克利勃纳提出了一种所谓三管齐下的研究策略。通过这种方法可以使我们在比较中来了解认知的发展。首先，应了解被试对于实验情况以及他是作为被试角色时的理解程度。从克派尔人的"聪明—愚笨"答案的例子中我们可以看到这一步骤是很关键的。的确，坎贝尔(Cornpell)认为如果我们没有从相反的角度获得令人信服的证据，我们就将会把在比较性研究中所发现的任何显著的差异都看作实验者与被试间交流的失败。第二个研究策略称为"具有实验的实验"(即，实验者应采用某种文化背景中某

一单一范例的多种变式，而不是在各种不同的文化背景中采用同一固定的范例）。这种方法的一个极好例证是科尔及其同事所做的关于克派尔人自由回忆的研究。第三个策略是找出一系列情境中的相同过程，这一系列情境包括自然出现的文化情境，以及实验或准实验环境。

根本的问题在于我们需要将实验研究和人类学方面的研究结合起来，以便在从自然发生的到实验的这一系列情境中找出某种特定的认知活动。我们认为，当组间差异表现在年龄上或在某种社会条件下的学业成功的程度上，而不是表现在其民族或正规学校教育程度上时，上述策略对于比较性研究是十分适合的。同样，我们认为，在对传统实验室任务上所反应出的组间差异作解释时，或为解释发育未全及其“正常”发育状况而需扩充智力测验项目时，同样都需要慎重。

B.社会历史环境：文化背景内的比较

科尔和斯克利勃纳在他们的研究工作中采用了跨文化的、心理人种史的方法。之所以采用这种方法，主要是由于他们考虑到要将心理过程从它所必须存在于其中的环境及内容中分离出来进行研究。这样做即使不是不可能的，也是相当困难的。他们指出：“作为要考察所有的心理机制的心理学家来说，他有必要先去考察一个以社会及文化背景提供材料而正常工作的机制。并且，即使在实验室所造成的人为的孤立环境中，他也无法摆脱现实生活内容的影响。同样，如果人类学家要了解现实生活的内容是怎样在历史发展过程中存在和发展变化的，那么，他就要懂得一个个体对文化背景所提供的材料进行的认知过程是怎样的。”

苏联心理学界在记忆发展的研究中是以类似的观点作为他们的基本理论的。他们在研究某一社会中的认知发展时强调社会历

史环境的因素。在最近关于苏联这方面研究的一份报告中，米歇姆(meacham)指出，在苏联的研究方法中包含着三个方面。我们在这里讨论其中的两个方面。第一，个体的动机和活动与内容之间存在着相互作用，正是由于这种相互作用才决定人们要记忆什么。第二，个体所表现出的特定的认知活动在很大程度上是由文化和历史条件所决定的。在这里，我们可以看到，苏联的这种辩证理论与科尔所采用的心理人种史之间在立场上是极为相似的。可能正是由于这种相似之处，使科尔对苏联心理学一直保持着兴趣。

通过这方面的讨论，使我们更加认识到，必须把社会一生态因素看成是在我们的社会中以及具有不同的文化传统的人群之间形成不同的认知活动的塑造者。另外，我们认为，为了了解幼儿的记忆能力及记忆方面的不足，我们有必要根据学前期自然出现的现象来考虑传统实验室任务的生态学意义。苏联心理学的“主导活动”概念与此看法是接近的。苏联心理学家认为，“主导活动”在认知发展的各个阶段都控制着认知活动的组织。他们认为个体的主导活动改变着个体的发育，而且主导活动又将其它的活动组织于自己的周围。主导活动的结果取决于成长中的儿童所处的特定的社会历史条件。苏联心理学对记忆的研究是把重点放在认知活动的发展上，而且认为认知活动的发展是受文化因素影响的。认为记忆服从于活动目的、服从于与文化因素有关的主导活动。他们从活动的角度对记忆进行了大量的研究。他们强调记忆的动机因素和文化历史条件，并且在方法上提倡采用观察及描述的手段，对实验控制则不太感兴趣。与此相反，科瓦勒(kvale)则认为记忆发展的非辩证研究是把重点放在无意义材料上的，并且忽略了记忆总是在各种环境中针对各种内容发生的这一事实。他认为，环境和内容有力地影响着记忆的方式。

综上所述，认知活动是在社会历史文化背景下发展和变化的，而对文化适应过程的性质影响着个体的主导活动、动机及认识活动的类型。所以，把发展中的儿童的记忆能力与儿童期的生态学结合起来进行研究是很有意义的。另外，为了评价在发展过程的各个阶段中个体典型的发展过程类型，我们有必要考虑影响其发展的其它诸因素，而不要把眼光仅局限于个体在人为创造的实验任务或智力测验的成绩上。至于兴趣的特定作用，则应在不同的情况下予以考虑，包括儿童期自然出现的主导活动。

C.作为一种文化背景的学校

因为正规的学校教育对于认知的发展有如此之大的影响，以至于会使人们把它认为是一种文化背景的缩影，在这种缩小了的文化背景下，人们强调某些专门的技能，并使这些技能精益求精。记忆的技巧及质量所受到的学校经验的影响丝毫不亚于所受到的其它认知活动的影响。正是在正规的学校教育的条件下，我们曾讨论过的那些变化才得以发生，例如各种可用的记忆方法的变化，个人做为一个学校教材的记忆者，对于自己的胜任能力和易犯错误的了解上所发生的变化，这没有什么可奇怪的，因为学校里进行着现代社会的主要的文化教学活动。在现代社会中，记忆是那些专门化的技能所特有的活动。无论是学习还是运用这些技能，都需要有记忆。在学校教育环境以外，在那些未受过学校教育的人及学前儿童中，这些活动是不常有的。作为一个学校，它极力鼓励的活动就是进行有意记忆，而不希望把这种记忆变成达到某一有意义的目标而采用的方法。

所以，人们可以期望正规的学校教育能总结出培养某些专门技能的规律并能改进这些技能，以便满足学术研究的特殊要求。这里所讨论的类似的记忆策略是特定的文化因素影响的结果；是

学校教育的结果，而不是人类成熟所带来的结果。对周围环境的适应过程影响着认知发展的方向，而学校教育仅是这诸多环境之中的一个。而且，用历史的眼光来看，对学校教育这一环境的适应只是人的认知发展过程中的一个新近的发展。在学校教育环境之外，其它许多完善的认知活动都可以带上认知发展过程的特征。

下面，我们来看一个关于依赖于外部援助而进行记忆的"自然发生"策略的例子。弗雷沃及他的同事们提供了许多关于幼儿根据自己的喜好而运用外部工具来记忆现实生活中的信息的例子。他们甚至在记电话号码这样的材料时也是利用外部存贮(抄写)而不是运用内部存贮(背诵)的。弗雷沃指出："在现实的、非实验室的生活环境中，人们大量地使用外部存贮及检索的办法。人类是这样，动物也是这样。在现实生活中，人们利用事物作记录并记录下事物，人们利用书籍、录音机、录象机、电影及计算机等可以进行极大容量和无遗漏的记忆。还通过别人的帮助来存贮和检索外部和内部信息。现实生活中的记忆任务具有开卷考试那样的性质，尽管记忆的研究者的任务并非如此。"

学前儿童是很少遇到需要对孤立的材料进行周密细致的内部记忆的情境的，而传统社会的人也不愿意在日常生活中遇到这样的情境。而且，我们社会里的成人在脱离了学校教育环境后，也很少进行这种如此深奥的心理活动，只是进行一些机械记忆，比如对某些数字的记忆(如电话号码等)。斯克利勃纳则认为，那种流行的自由回忆法对于跨文化的比较是很不适用的，因为它没有提供回忆的外部线索，而要求人们寻找内部线索以组织操作。她认为，对于内部的，重新提取的线索的过分强调是不正确的，因为这样会严重脱离自然出现的及日常生活中所遇到的记忆问题的实际情况。她还引用了马格利特·米德(Margaret Mead)早先所进行的关于原始人依赖外部线索进行记忆的情况的观察实验。米德指

出："南澳大利亚的土著居民有一种特有的记忆方法：为了能够正确地讲述很长的图腾神话故事，他们就必须到神话故事中提到的那些地方去散步，以便记住这些情节。提醒他们记住正确情节的刺激都是他们生活环境中的一些外部事物。"这种方法和著名的轨迹法有惊人的相似之处。我们可以回想一下轨迹法是怎么产生的。这种方法是为了适应罗马的演说家们的需要而产生的。这些演说家们为了作好演说，必须采用某种技巧来把演说中的主要材料的排列顺序牢牢记住。我们知道，事先不需要什么学习，幼儿和成人就都能有效地运用这种技巧来进行记忆。这是对依据外部线索可以有效地进行记忆的一个强有力的证明。

对于地点、环境以及其它物理线索(如符号等)的运用可以帮助记忆，这种运用是依据线索而进行的记忆的一种意义化过程。在这一过程中，所要记忆的材料被置于现实生活的经验之中。我们认为这种记忆形式是人类记忆中的一种占优势的形式。为了记住一些人为编制的材料，依赖孤立的内部线索而进行的记忆是人们认知活动的一种专门形式。这种记忆不仅是典型的欧美文化传统影响的反映，而且也可能是正规学校教育本身的一种特有的产物。这种记忆技巧并不一定代表认知活动的基本过程。

我们认为，以某些记忆形式为基础的人类概念的发展存在着基本的普遍性和连贯性。这里所说的某些记忆形式就是纳尔逊曾提出的将记忆材料情节化、连贯化的记忆形式，他认为，这种情节化、连贯化的过程就是将围绕某一行动或某一目的所要用到的各种不同的概念有联系地组织起来的过程。他提出的这种记忆形式与科尔及斯克利勃纳所提出的与个人经验的背景及事件有关的自然记忆类型是很相似的。科尔和斯克利勃纳认为个人经验的背景及事件使人具有对周围事物的组织能力及预测能力。纳尔逊所提出的记忆类型与哈夫洛克(Havelock)所提出的在口语传统条件

下形成的记忆类型也是很相似的。因为海夫洛克认为在口语传统条件下，人们的注意力是集中于在有意义的情境中行动着的人以及所发生的事件上的。

纳尔逊指出，对现实环境中所发生的事件进行组织，并使之情节化和连贯化是幼儿认知活动的重要方式。他认为要进行这一系列认知活动就需要有实际的经验，并要求在时间及空间上具有有意义的重复的结果。这对幼儿来说是一个很重要的组织过程。我们相信，无论是对我们社会中的成年人还是对其他社会的未受过学校教育的青年来说，这都是一种占优势的认知形式。另外，在正规学校教育环境中还可以看到，记忆受到经验的新近性和程度的影响。在这种情况下，人的思维更多地受到对现实生活中的活动的组织的影响。虽然对周围事物在记忆时加以组织使之意义化等一系列过程是受环境制约的，是根据所发生的事件进行的，但对于日常的认识活动来说，这一过程又具有普遍性。通过那些特定的，重复性的经验我们可以抽象出它的一般原则，并依此去解释新事物，预测其相似性，提供一个关于全人类的组织结构。

总之，我们可以认为，在人类概念的发展上存在着根本的普遍性。而且认识活动的基本方式就是对周围丰富事物的认识结构的形成和完善，其重点在于那些重新出现于个人经验中的行动结果。后来出现的、与环境相对独立的结构或规则只是基本的认识活动的补充、完善，而不是代替。尤其值得注意的是，对于不同的主导活动背景，人们发展了适用于不同记忆类型的专门技能。例如，在我们社会的学校中记忆课文的技巧、为表达史诗而采用的口头记忆技巧，或在传统社会中为了辩论而记忆神话人物的名字。这里我们可以看到一个值得注意的事实，即在西方社会，尽管没有精心地指导，然而对正规学校教育的向往就是决定记忆发展的方向。理解和保持信息的活动是学校教育期间的主导活动，而理

解能力与保持是指理解和保持课文中所出现的孤立的、公式化的语言所包含的信息。所以，为达到这一目的而进行的专门的认知活动的发展应被置于发达的现代社会中的学校教育这样的历史环境中来进行讨论，而不应把它看作是人类认知发展的自然的最终产物。

D.学校与条件不利的儿童

每一种文化背景都有它自己的方法来提高基本的认知能力，以满足特定的需要。在我们的社会中，在学校里，尤其是高年级，教学工作的重点是放在与环境相独立的认知技能的培养上的。正如布鲁纳及奥尔逊指出的那样："无论是为了经济上的原因还是为了效率上的原因，学校都独立于活动环境之外而进行教学，这样的教学是以与环境相独立的、基本上是象征性的各种手段来进行的。"他们还认为："由于学校里的这种与现实环境相独立的气氛的持续影响，学生们开始根据课文中的内在联系而把课文所叙述的东西认为是真实的。这不仅是因为从经验角度上看是可能的、或者是他们所经历过的，而且因为这是课文里其它部分的必然延续。在学校中，即使不能全部地实现这一点，最起码也应作为一种指导思想来贯彻。"

尽管人们可以论证这一假设，即西方出身于中产阶级的大学生的思维方式应该成为我们所有社会成员认知发展的目标，然而，毫无疑问，学业上的成就有赖于在这一领域中是否能达到合适的等级，事业方面的成就也在一定程度上依赖于此因素。人们希望儿童从与环境紧密联系的、以经验为基础、以游戏活动为中心的学前生活转入与环境相独立的、非个人的、为学习而学习的学校生活中去。一些孩子完成了这一转变，而有的儿童却没有完成，根据儿童学前期的准备活动来看，许多儿童不应被视为智力落后。

如果他们的这一转变是由于其它因素的影响而没能实现，那么我们不能认为他们智力落后。

让我们简单地分析一下那些使条件不利的儿童的转变比中产阶级儿童更加困难的各种因素。我们抛开如学习一种新的语言（无论是学不同的方言或是学第二门语言 ）这些明显的问题也会发现条件不利的儿童因严格的语言规则而感到困难重重。这种严格的语言规则有利于语言的上下文有关的、社会以及修辞的功能，而对于学校里所要求的上下文无关的、明确的交流模式却并不那么有利。另外，我们也知道，条件不利的儿童在学校生活的初期很难解决那些脱离开经验支柱的问题，而且也很难采用适当的记忆方法及搜索策略来促进学习。他们在发展这些技能时要得到明确的指导，甚至不能把这种训练效果推广到类似的活动中去。或许是因为这一问题，他们的元认知发展受到了阻碍。简单地说，儿童如果在问题解决活动方面没有足够的成功经验，则很难指望他能对自己的认知活动进行有效的管理。

也许有一个更为基本的问题。早期失败的经验会严重地损害儿童的自我概念。他们会不相信自己还能作为对要学习的东西进行探索的活动中的一员。如果他们没有求上进的愿望，那么他们在学业上将肯定是无所造就的。儿童在很小的时候就会出现这种学习无助现象，但这种现象并不是天生就有的，是在后天生活中形成的。儿童对于自己的认知过程的认识肯定会受到学习中竞争的感受的影响。许多条件不利的儿童对学校里的学习竞争不怎么感兴趣，尤其是那些被视为无竞争能力因面选去接受特殊教育的儿童。

布鲁纳曾经指出，学校作为一种公共机构，脱离了适合于儿童的早期游戏活动的思想，甚至也脱离了适合于儿童的早期游戏活动的思想，甚至也脱离了成人所要求的职业活动。例如，在原

始社会，儿童靠模仿成人的示范来进行学习，这种学习是与游戏交织在一起进行的(如狩猎、纺织等)。他们从游戏活动向真正的成人职业(即从玩狩猎游戏到真正狩猎)的转变是逐渐实现的，儿童时期的探索性的游戏活动与成人的日常事务间并不存在明显的界限。在我们的社会中，学校充当了这两方面活动的调解者，但学校却又未能确定这两种活动间的必要联系。这不仅是因为学校所采取的是一种脱离生活的过程(学校不鼓励游戏活动，并且也不鼓励通过活动的方式来学习，面只鼓励通过听和阅读所进行的学习活动)。但是，通向成人社会的这种联系究竟是什么并不很清楚。布鲁纳指出：学校就是一个被它本身所进行的那种难懂的工作所隔离了的世界，正如麦克鲁汉指出的那样，学校是一个媒介，有它自己的交流方式，而不管所教的是什么。这种交流方式与工作，与成人生活是无关的。对于那些为了自己而渴求知识的人来说，没有什么令人不安的。而对于那些不想学或不能学的人来说，学校却不给予任何指导——仅仅是知识，这些知识与生活的关系何在，学生和教师都不甚明了。学校中有某些脱离生活和导致混乱的因素。

如果学校的教学活动不与现实生活中儿童每天都遇到的游戏及工作的经验联系起来，那么注重事业的人就会认为课堂教学没有什么意义，这是不足为怪的。如果课程没有意义，那么为什么儿童还根据那些似乎合理的或微妙的标准来检查自己的成绩呢？这种学校与现实生活相脱离的问题甚至在霍尔特所说的那些中产阶级、成绩优良的学生中也能看到(参见第三节)。但主要的问题在于条件不利的儿童，对他们来说，要适应学校生活就必须使学前生活条件有一个根本的转变。除非给这些儿童以帮助，否则，他们是学不会进行学校式的游戏，学不会作为学生所应掌握的技能的。

我们还可以看到这样的事实，向学校教育生活过渡的困难程度，对不同的儿童来说是不相同的。有的儿童根本没有进行这样的过渡，有的儿童则以失败而结束这种过渡。这是否是由于生物学或社会经济学还是某些因素的某种结合所致？另一个问题是，限定了许多教学实践的神秘性之后，我们是否希望所有的孩子都遵守这些标准呢？尽管回答是否定的，但是赞成把那些能进入高校就学的人与不能进入高校的人区分开来的观点引起了广泛的争论，这里不再论及。

另外，我们认为，为使尽量多的儿童根据他们的传统技能来培养他们的能力是非常有益的。所以，应对所有需要帮助的儿童提供指导。学校之所以能存在、能发展到现今的水平，全在于它满足了现代社会科学技术发展的需要。因此，学校的教学活动未必会有根本的变化。这样，我们必须为尽量多的儿童做好准备工作，以适应学校的要求。这意味着为那些在向学校生活的转变中有困难的儿童而改革教学方法，舍得在他们身上花气力。

正因为这些理由，我们对可教育的智力迟钝儿童越来越关心了。而且最近我们对那些可能被视为智力迟钝的儿童也更为关心了。许多可教育儿童都深受"学校病"之苦，在学校里仅仅是因为他们智力落后及反应迟钝被正式地挑选出来。而在上学之前，他们要么没有体验过这样的困难，要么这样的困难没有表现出来。毕业后，他们中的大部分人又能为社会所接纳，仅仅是在校学习期间他们才受到一种无形的隔离。如果他们能够得到帮助，并由此达到学校的一般要求，那么这个好处是非常大的。他们将不会被分到进行"特殊教育"的分校去，而随之而来的则是他们将会丧失自信心。

有一种无疑是不完善的理论经常出现并影响着人们对那些挑选出来的落后儿童组成的特别班级的指导工作。这些儿童常被刻-

上这样一种印记，即他们学东西太慢，而且非常难教。但这些结论常常是根据从孤立的实验室研究或IQ测验所得的资料的分析中得出的。对这些研究工作或测验中所使用的材料及程序，孩子们是不熟悉的。我们倒宁愿持这种观点：在对于文化背景内的比较性资料的解释时，科尔及斯克利勃纳对认知过程的跨文化比较工作所提出的注意事项应该得到更进一步的重视。在实验中，被试没有归纳出某种累积的复述法，并不一定意味着他不能归纳出这一方法，他们明显地能够，而且也的确在将知识灵活地运用于现实生活中去。跨文化研究中很需要对某一特定的现象进行多重的观察。任何旨在对年龄、民族、道德、IQ分数等方面有差别的各组间进行比较的研究工作都不可缺少这种多重的观察。所以，我们坚决支持科尔和斯克利勃纳关于比较性研究的三点方案：①弄清被试对实验情况的了解程度；②进行"实验的实验"；③弄清在一系列情境下(自然出现的情境，准实验情境及实验情境)的相同过程。我们还不曾听说有哪些关于学习能力低下儿童认知发展的研究工作达到了这些标准。

既不是为了理解也不是为了鉴赏的目的而去学习孤立的材料，这对具有一定技能的人来说并不是一件容易的事，而对于那些生手来说就更加困难了。如果学习慢的儿童认识到这种困难的学习是学校里的主导活动，那么他就注定要失败。所以，学校应采取措施使学校的早期活动仍与现实生活紧密联系，以便帮助那些学习迟钝的儿童解决他们的适应方面的困难。

前面讨论过的关于口语和文字材料传统间的区别将会有助于我们找到一种新的途径来解决条件不利的儿童在学校教育生活中所遇到的困难。我们不应该认为条件不利的儿童缺乏某些技能，这种新的方法能使我们发现把在口语传统下条件不利的儿童所掌握的技能作为我们的研究重点是不无裨益的。如果我们在寻求那

些可以被学校利用的能力，以便更好地培养落后儿童的认知发展的话，要做到这一点，那么从口语传统背景下儿童所具有的能力上来开始我们的研究工作是非常合适的。可教育的智力迟钝儿童很适合于记忆他们在日常生活经验中所遇到过的地点、人物等。他们对于家庭、街道等环境的情况的记忆是很好的。他们对民间歌曲中的词句、垒球赛的比分、电视节目的时间等有惊人的记忆力。尤其是对民歌的记忆可以保持相当长的时间。所以，我们可以认为，如果能做到以下几点，我们就可以使条件不利的儿童能够比较容易地向正规学校教育生活过渡：①研究工作的重点应放在口语传统背景下所具有的技能的潜力上，而不应放在他们尚未获得的文字材料传统背景下所应具有的技能上；②训练期(观察与尝试)的游戏式学习期限应尽可能长一些；③根据活动与儿童对于自然出现的活动所具有的认识之间的联系程度和意义对想要进行的任何活动的内容及环境进行一番细致的分析。

提出这样的主张是容易的，但要做到就困难了。然而，如果这种主张能够严格地实现，那么对于人们实现这一程序的具体过程应该详细地记录下来，这是最起码的要求。从实验心理学家的眼光来看，详细地叙述研究过程的每一步骤将会给我们提供必需的例证。在这里，我们试图做到这一点。然而，我们还要强调，所有的研究工作都有一个根本的指导思想，即提供有实用价值的信息。这就是说，这些研究工作应该而且也能够在教学情境中得到充实。

随着时间的推移，训练可教育儿童的记忆方法及元记忆技能方面的设想也在发展。我们来看一看这一发展的情况。我们知道，培养某种传统的记忆程序的方法不会有多大的发展，这样的方法是不足取的。因为这种方法的受益范围很窄。应在有意义的环境中训练儿童的记忆技能并同时明确地解释为什么这些技能是有目

的的。这种做法则是可取的。当然，这也是我们实验室普遍采用的方法。(见第四节)

在方法上的一个更加根本的变化就是，我们认为对于作为训练中心问题的技能应给予极大的重视。我们可以选择某种类型的认知活动来开展我们的训练工作(所选择的类型应能在训练中给落后儿童以强有力的帮助)。这种认知活动的类型应具有以下几个特点：①在情境变换时，应能表现出很大的适用性；②能使儿童容易地感觉到这样的活动是合理的，是可行的；③应与现实生活的经验有某些共同之处；④这一活动所包含的各个过程应很好地理解，以便找出有效的训练技巧。我们倾向于选择一些元认知活动的子系列。我们认为这些子系列能满足提出要求、检查核对、观察并进行真实性测验等要求。当然，这些工作仍然还有不明确的地方，我们打算挑选出一些基本的技能，以便进行深入细致的研究。我们所选择的认知活动的类型可以归于一个总的题目之内，即"自我讯问"。

从苏格拉底式教学法的意义上来说，我们的最终目的就是训练儿童去辩证地思维。苏格拉底的方法中，教师不断对学生的假设和前提提出问题。教师扮演明知不对而坚持错误观点争论不休的人，并且探查学生的薄弱环节，以此方法从反面来教导学生，这种做法的结果是比较好的。这样做的结果能使学生通过自我讯问来执行教师的职能。尽管科林斯描述的这种技能显然不能直接运用于年幼、学习能力低下的孩子，但是这种方法的基本原则是适用的。在面临某些问题(如教学上的问题、教学问题、实验室任务)时，我们应从一个非常低的水平开始教儿童自我提问。切实可行的自我提问的问题类型可以和事先自我提问的一套n个问题类似。例如：①停下来想一想；②我知道该干什么吗(即理解那些明确的或含蓄的指导语吗)？③在我开始以前，还有什么需要知道的吗？④我

是不是已经知道了那些有助于我的内容了呢(即：这是一个与我以前解决过的问题相类似的问题吗)？

我们打算普遍训练可教育的儿童遵照口头的或书面的指导语来完成种种简单散文的理解任务，这些作业都置于有意义活动的环境之中，好象装配一个玩具或按处方配药一样。在这些活动过程中，他们就象上面所说的那样进行非常认真的、显然合格的自我讯问。我们相信，制订简单的系统来启发孩子们的自我意识，并引导他们通过自己的活动进行意识调节是训练工作的重要形式。因为这样的最终结果是令人满意的，这种结果是靠自己的力量而产生的，一定能适用于不同的情境。而且这种结果还一定能促进儿童认知和元认知技能的发展，以及促进对个人能力的了解。

Ⅵ. 小　　结

这一章的主题是某些元认知技能的发展问题。这些元认知技能是在实验、教育以及自然出现的各种情境下有效地解决问题的标志。我们认为，知识与对这些知识的理解之间是有区别的。这种区别既重要，又有效。可以给关心认知发展的人以极大的启发。这一章的重点在于探讨某些儿童认知结果的内部执行过程，现代认知理论中所讲的执行过程是指预报、计划、核对及监控。我们认为，这些执行过程是在广泛的学习环境中进行有效思维的基本特征。

目前所达到的工艺水平里，大多数发展和训练研究都涉及到了对几种简单的记忆技巧进行有意识的控制，以便更好地进行有意记忆的问题。这也反映了我们的研究艺术，既从跨文化的生态学标准的角度，又参考学校在教育落后儿童时所遇到的一些问题来讨论这些技能的文化相对性。

本章的一个重要的目的就是强调指出，超脱开传统的记忆任务框架之外的一般元认知技能的实验研究还为数不多。应制定出成长中的年轻人的有效的训练大纲。这种大纲的重点要放在执行机能上，而不应放在某种技能的完善上。但是，这方面的工作却是一个被忽视了的研究领域。旨在启发那些不能自发地运用简单的核查技能的人的训练方法至少在学校学习环境中或传统实验室任务中没有得到发展。尽管我们不应低估在制订这样的训练措施方面所遇到的问题，但是，只要我们努力地工作，发挥我们的独创性，那么无论在理论上还是在实践上，我们都会有越来越大的收获。

孟鸿伟　赵尚武 译

三、皮亚杰学派和新皮亚杰学派：教学发展理论和教学工艺学的探索

罗 比·凯 斯

I. 引 言

正如许多学者曾经指出的，在学习理论与教学理论之间存在着一个重要的差别〔奥苏伯尔(Ausubel)，1963年；布鲁纳，1966年；格拉塞，1976年〕。学习理论是要提供学习过程的模型，而教学理论则旨在提供一个如何使学习过程最佳化以达到特定目标的模型。某一学习理论是否能为建立教学理论而提供一个可贵的起点，就依这类重要的目标而定。不过，某一学习理论无论怎样贴切，在某种意义上说，教学理论总是会超过它的，因为数学理论要为在特定的教育环境中作出正确的教学决策提供一整套原则。

在智力发展理论与教学发展理论的差别上，同样如此。智力发展理论的目的就是要为人的智力从出生到成熟如何变化提供模型，而教学发展理论的目的则是要提供一个如何使这个过程最佳化以达到特定的目标模型。某一发展理论是否能为建立教学发展理论提供一个良好的起点，就依这类重要目标而定。但是，某一智力发展理论无论怎样贴切，教学发展理论在某种意义上总是会超

过它的，因为教学发展理论要为特定发展阶段上的特定类型的儿童作出正确的教学决策而提供一整套原则。

本章首先简要介绍在智力发展的研究领域中占统治地位的心理学理论——让·皮亚杰(Jean Piagt)的理论。并以此作为一个起点，说明教学发展理论最终总是能解决些什么教学问题。在概略地叙述了这些问题之后，我们将介绍一个将皮亚杰理论作了进一步发展的理论。这个新理论试图为皮亚杰所描述的那些发展能力的获得过程提供更为详细的模型。随后，对有关这一新理论的研究进行评论，对前边提出的一些教学问题重新加以考虑，并打算阐明这一新理论对这些教学问题是如何作出回答的。最后，要提出若干符合新理论的教学原则和教学方法，并对这些教学原则和教学方法如何运用于课堂情境之中做一点说明。

Ⅱ. 皮亚杰的智力发展理论

皮亚杰理论的目的是要描述并解释人的知识的发展，他主要考虑的是人的知识的组织或结构，而很少考虑人的知识的内容。皮亚杰理论的主要假设是：人的知识是有结构的；之所以如此，是因为知识是通过活动取得的，而活动又是有高度结构的。皮亚杰研究的中心点就是探讨获得知识的活动的结构。

首先要考虑的是幼儿出生后的头两年获得知识的方式（皮亚杰，1951年，1954年）。当一个婴儿出生的时候，他就具备一些基本的动作图式。如：抓握、吮吸和视觉追踪等，儿童正是通过这些图式的练习来获得有关周围世界的最初的知识。在儿童获得这些最初知识的同时他的图式本身也发生了两种重要的变化：第一种变化是，图式逐渐变得更加分化了，这就是说，图式逐步变得能够细致有别地适用于品种日益繁多的物体；第二种变化是，图式逐渐变

得更加协调。例如，刚开始抓握是一种比较孤立的和反射性的图式，但是后来它逐渐变得与其它许多图式（如视觉探究，伸手和挥动手等）协调起来。由于这些感觉运动图式的运用和协调一致，婴儿不仅能对切近自身的环境中的物体得到多方面的和稳定的了解，而且还能获得一个灵活的感觉运动运算系统。婴儿用这种运算系统，能够对新环境中的新物体取得更进一步的了解。感觉运动运算系统的基本特点是：一种图式(如抓握)产生的任何作用都能被另一种图式(如移位)所逆转。

在幼儿的随后的六年中，在较高的水平上出现了类似的趋势(英海尔德和皮亚杰，1964年；皮亚杰，1952年，1960年)。由于儿童已经掌握了与他的切身环境有关的基本的感觉运动运算，儿童开始用符号表示这种运算，对这些物体除了能进行身体上的操作之外，还能进行心理上的操作。儿童在这个时期所发展的新图式同出生头两年所发展的图式的相似处在于：新图式仍然都只能应用于面前的具体物体。而它们的不同处在于，新图式现在包含的动作可以认为具有更多的内在性质。例如，组词和用词的程序、物体分类的程序和数量化的程序。起初，新形成的符号图式还是相当笼统的和孤立的。例如，儿童从两个物体中找出较大的物体这种心理活动还是以物体笼统的外貌为根据的，还不能同从两个物体中找出较小物体这种互逆活动协调一致起来。然而，随着时间的推移，这些图式再一次变得更加分化和协调了，使得儿童到6—8岁时开始形成新的系统。这样，在2—8岁期间，儿童不仅能获得自己环境中有关物体事物的基本类别和关系的知识，而且还能获得一个具体运算系统。儿童凭借这种系统，能在新环境中获得这类更深一层的知识。如同感觉运动的运算系统一样，这种具体运算系统的基本特点仍在于：任何一种心理活动(如，宽度的增加)的作用都能被另外某一种心理活动(如，宽度的减少，高度

的增加)所逆转或补偿。

从8岁到12岁，儿童的具体运算系统的稳定性和概括性不断增加。同时，这为形成更高一级的心理运算打下了基础(英海尔德和皮亚杰，1958年)。这些更高一级的心理运算是把第二阶段（心理运算)的成果作为构件并直接对它们进行运算的，而不是对具体的特征进行运算的。例如，在具体运算阶段，儿童亲眼看到天平臂的一端重量的增加，就能想象出用减少同样重要的方式来简单否定这种作用。同时，儿童也能理解在天平臂的另一端增加重量所产生的互逆作用。而处在下一个发展阶段的儿童，就能够把否定的运算或互逆的运算作为一个起点；在没有任何具体事物出现的情况下产生另一种心理运算。他们能够把这两种关系之间的关系作为另一个已知条件，进而想象出另外某种变量(如离支点的距离)的关系之间的一种反比关系。当儿童对各种运算加以运用和比较的时候，他们就能够在头脑中产生许多组合，这比实际出现在他们面前的组合要多得多。随着时间的推移，儿童发展了一个在给定的情境中产生所有可能的运算组合的系统；同时儿童还发展了一种用实验的方法从众多组合中抽取一个组合而排除其它所有组合，以便分离出某个特定变量的作用的程序。青少年除了对更加抽象而复杂的世界获得了解以外，还获得了一个协调的形式运算系统，凭借这个系统，他们能够在新的内容领域中获得同样抽象的知识。

那么，在这三个发展阶段中，自始至终影响儿童智力发展的各种变量是什么呢？皮亚杰认为有四种基本的因素，即：成熟、物理经验、社会经验和平衡(皮亚杰，1964年)。其中，他认为第四个因素是最重要的。尽管前三种因素的必要性是大家公认的，但这些因素产生作用的确切机制大部分还是不明确的，这就使它们的重要性随之而降低了(皮亚杰，1964年，1970年)。遗憾的是，皮亚杰本

人对平衡过程是什么这个问题也是相当含糊不清的(皮亚杰,1964年,1970年,1971年)。从全局的分析水平上说，皮亚杰清楚平衡过程包含着对现存一整套运算的适宜性的反思以及对新的运算的实验。另外，皮亚杰也清楚儿童确定运算适宜性的标准，这就是一致性——预见与实际观察到的事物的一致性。在一种情境中所作的判断与在另一种情境中所作的判断的一致性；或者根据一个运算系统所作的推理与根据另一个运算系统所作的推理的一致性(皮亚杰,1971年)。但除此以外，皮亚杰并没有阐明平衡过程发生的确切机制。

虽然皮亚杰的理论并没有对平衡过程进行详尽的阐明，但是皮亚杰所做的描述足以得出结论说：这个发展过程必然是一个相当缓慢的过程。正如人们不能奢望用激素治疗促进成熟或者用填塞多种多样的新物理经验和社会经验促进学习的办法来促进儿童在几个月之内就可以成长成人；同样，人们也不能奢望简单地依靠成熟或学习的作用，使儿童在几个月之内形成一个形式思维的完整系统。由于平衡过程是一个包含有反思、协调和构造等活动的内部过程，又由于认知建构任务是一个巨大的、基本的任务，皮亚杰乃假定：期望通过任何一种环境的管理或控制使得儿童加速发展，其发展量是有一个限度的。

总之，据皮亚杰看来，儿童出生时就已经具备了一种动作图式的原始指令系统，通过平衡过程——在某种程度上借助于成熟和由外部因素所引起的学习——这种动作图式的初始指令系统逐渐变得更加分化和协调。在儿童生活的许多方面，这些已经协调了的图式就具有一种相互依存性，这种相互依存性又使这些图式获得了那种为一个有组织的系统或运算结构所必须具备的特性。智力发展就是由一系列这样的结构所组成的：开始是感觉运动期的运算结构，最后发展到形式思维期的运算结构。

Ⅲ. 教学发展理论的几个问题

如果皮亚杰对智力发展的论述在总的轮廓上是正确的（我也认为是这样），那么，任何一种理论想要表明怎样才能最恰当地促进儿童智力发展，定要至少解决两个问题：第一个问题是如何使教育环境具有良好的结构，以便使智力运算主要系统的获得最佳化。这个问题是以下面两个假定为前提的：(1)这些系统的获得能够受到各种潜在地为人所控制的环境因素的影响；(2)从发展的观点来看，儿童当前的教养环境已经不是最理想的了。看来，这两个假定，是有根据的，它们有充分的根据证明，皮亚杰所说的各个阶段的发展速度，是千差万别的，它是儿童所处的人的环境的函数〔达森(Dasen)，1972；高迪亚(Gandia)，1972；霍洛斯(Hollos)，1975〕。另外，还有一个充分的根据是，许多（不是大多数）成人在现存的环境条件下从未达到形式运算的阶段〔卡普卢斯(Karplus)和皮特森(Peterson)，1970；库恩·兰格(Langer)，科尔伯格(Kohlberg)和哈恩(Haan)，1977；皮亚杰，1972；托姆林森-基赛(Tomlinson Keasey)，1972〕。

要解决的第二个问题是，文化上有价值的事实和技能的教学应如何设计，以适合学习者已经掌握的知识—采集(Knowledge-gathering)运算系统。这个问题是以下面的假定为前提的，即不适合于学生可用的运算系统的那种教学不是最佳的。再一次看到，这个假定是有根据的。当用比儿童自己原来用的更高的思维水平对儿童进行解释或演示时，甚至在几秒钟之后儿童也不能把这些内容重现出来。相反地，他们甚至还会以适合自己思维发展水平的形式曲解别人给他们解释或演示的材料〔布拉特(Blatt)和科尔伯格，1971；图里尔(Turiel)，1972〕。在上述两个问题中任何一个问

题能够解决之前，仍然还有一些理论问题和技术问题留待解决。本节的目标就是要探讨这些理论问题和技术问题的。

A.使运算结构的获得最佳化

首先要考虑的一个问题是如何使环境具有良好的结构以便使运算结构的发展最佳化。从全局的分析水平上看，为了达成上述目标，提出一个皮亚杰学派的方法，是没有多大困难的。第一步是详细说明主要的运算结构在数学、历史、自然科学等等这些对教育者来说极为重要的领域中的表现方式；第二步是确定受教育的学生当时正在起作用的一般运算阶段；第三步是使课程具有良好的结构以便包含能促进学生思维发展从当前的水平过渡到高一级水平的练习。虽然促进运算发展的一般程序是清楚的，但任何一个特定步骤真正付诸实践之前，还有着大量的工作要做。

1.第一步：结构分析。让我们把也许是皮亚杰最著名的作为物质守恒的基础的运算结构分析，作为对通常的学业任务结构分析所固有的问题范例〔英海尔德(Inhelder)、辛克莱(Sinclais)和鲍维特(Bovet)，1974年；皮亚杰1957年；皮亚杰英海尔德，1941年〕。这里所获得的特定的概念就是：尽管容器里的物质的形状看起来是改变了，但容器里的物质的数量仍然不变。如果在两个形状相同的量杯里盛两份水，并且建立了一个初始的等量，那么，当我们把一个杯子里的液体倒入另一个高而细的杯子里时，尚未获得守恒概念的儿童就会认为这两个杯子里的液体不再等量了。而已经获得守恒概念的儿童，则认为这两个杯子里的液体仍是等量的。他的解释是：什么也没有加进去，什么也没有拿走；两杯水的数量原来就是相等的。或者还可以换一种说法，他会简单地说：如果把细而高的杯子里的液体倒回到原来的杯子里，那就再一次看到这个杯子里的液体同标准杯子里的液体是等量的。

在皮亚杰看来，那些掌握了守恒原则的儿童之所以能做到这点，是因为他们有了如下的理解，即：在转换过程中所出现的显著维度量的增加是以不显著维度量的减少作为补偿的。这种认识可以用下面的符号表示：$a_1 \times b_1 = a_2 \times b_2 = a_3 \times b_3 = \cdots\cdots$其中a和b分别代表高维度量和横截面面积维度量，符号下角的数字表示不同的假设值。皮亚杰指出，这只有会进行逆运算的儿童，才能有这样的认识，他们才能想象出一个维度量的增加是由于另一个维度量的减少的结果。鉴于可逆性是具体运算阶段的标志，所以皮亚杰得出结论说，儿童获得了具体运算系统才能获得关于守恒概念的知识。

虽然这种分析或许符合皮亚杰的目的，也说明了把这种分析推广到课堂任务中的种种努力所存在的问题。第一，一组如上所述的逻辑运算，在除了也许是数学或自然科学之外的任何学科领域里，究竟会起什么作用，这是一点也不清楚的；第二，皮亚杰早先使用什么程序来揭示上述逻辑群的作用，这也是不清楚的。儿童完成守恒任务的实际作业并不表明他们在心理上正在进行一种高度差与横截面面积差之间的相互补偿的运算过程。很清楚，确定基本运算结构的程序，不仅仅是分析儿童对任务的反应，而必须要分析这种作业所要用到的知识。遗憾的是，皮亚杰对如何作出这种合理的分析并没有做出总的论述。因此，即使象在自然科学或数学这种虽然需要进行逻辑推理的完全确定的领域中，如何识别出特定任务所依据的运算结构，这是不清楚的。

2．第二步：个别评估。假设结构分析的问题能解决，那么设计一种发展性课程的下一个步骤就是确定接受教学的儿童当时起作用的一般运算水平。从表面上看，这个问题似乎价值不大，因为皮亚杰学派的若干套测验已经得到发展并标准化了〔戈德施米特(goldschmidt)和本特(Benteer)，1968年，图德纳姆(Tudden-

ham)，1970年；温克尔曼(Winkelmann)1975年〕，这些测验只是施测的问题了，而且也指明了根据学生的运算水平决定使用什么样的量表。事实上，问题并不那么简单。

首先，如果检验一下那些被假定能测选十分相似的结构的皮亚杰学派的测验各项目之间的相关，那么人们就会发现，这些相关在正常情况下是很低的，有时甚至是不显著的。尽管在相同心理测验的两个测验组之间显示出一个低的交替相关并不是不常见的，但是，从教学的观点来看，我们可以从中得到启示；受教儿童当时起作用的一般运算水平，必须从既定任务所要求的特定结构方面来加以评估。即使是有关结构测验之间的低相关是由一些非控制任务因素的作用造成的，这也不可能知道某个儿童的作业究竟是由于具有或者缺乏起作用的特定运算结构造成的，还是由于对一个非控制任务因素的反应造成的。

其次，如果再看一下皮亚杰学派的相同运算结构的测验，就会发现一个甚至更为严重的问题。尽管项目之间相关相当高的特定的作用，而成功的绝对水平则是千差万别的。这些水平随着所采用的特定的测验项目的不同而大不相同。例如，如果通过数的守恒测验的年龄大约是6岁，那么通过重量守恒测验的年龄大约是9岁或10岁，而通过体积移位守恒测验的年龄大约是11岁。从发展理论的观点来看，在用来识别基础结构的任务中划定这种"儿童发展阶段"("decalage")，其严肃性还是一个值得争论的问题〔贝林(Beilin)，1969年；帕斯考尔-莱昂(Pascual-leone)，1972年；皮纳德(Pinard)，1975年〕。然而，从教育评估的观点来看，这个问题的严重性，是不可能有争论的。如果某一既定的结构幼儿园的年龄就已经学得，却到了中学年龄才用到某些任务上去，那么，去了解儿童是否获得这种结构，就毫无用处了。只有了解到儿童是否能把这种结构运用到与教学有关的那种任务中去，这才是有用

处的。然而遗憾的是，那些影响智力结构运用于特定任务中去的因素，皮亚杰的理论根本没有详细加以阐明。

3.第三步：教学设计。假设结构分析和评估的问题能得到解决，那么设计一种发展性课程的最后一步中就是要设计出若干练习以促使儿童从现有的功能水平向下一个更高的功能水平过渡。正如上面第二部分所说，皮亚杰对这种阶段一过渡的过程中所包含的内容是很不清楚的。让我们来看看皮亚杰对儿童如何获得作为掌握守恒概念的基础的那种逻辑结构的描述吧。皮亚杰把这个过程分为几个亚阶段(皮亚杰，1957年)：第一步，儿童主要根据高度估计两个杯子里做比较的水的数量；第二步，儿童主要根据横截面的面积估计两个杯子里做比较的水的数量；第三步，儿童进行单维度量估计的两种模式发生了冲突；第四步，儿童用通过把两种较低级的模式协调起来建构一种较高级运算结构解决了上面的冲突；最后，第五步，儿童巩固和扩展这种较高级的结构。

作为儿童在获得估计数量的复杂结构时所经历阶段的一个模型，上面的这种描述对我来说似乎是很令人信服的。它也得到了一些实验的支持。利菲弗里(Lefefvre)和皮纳德(1972年)在最近的一项研究中，精心设计了一套教学活动以促使儿童通过上述的各个亚阶段。结果，他们使儿童对守恒的理解发生了很明显的变化。应当指出的是，虽然皮亚杰对儿童所经历的各个亚阶段做了适当的探讨，并且也弄清楚了激励这种进步的一般因素是儿童对连贯性的探索，但是，皮亚杰没有详细阐明从一个亚阶段向另一个亚阶段的过渡。他恰恰没有说明儿童察觉出或解决高度和横截面面积之间的矛盾时所经历的内部过程，或者说没有对那些影响这个过程的具体的成熟方面的因素和经验方面的因素作任何说明。

从发展理论的观点来看，对象这样一种说明的重要性是会提

出疑问的。可以辩解说，至少在理论建立的早期阶段，只要注意起作用的一系列一般因素(这里是指：成熟、经验和自我调节活动)，并且描述这些一般因素之间的相互关系就足够了。而从教学的观点来看，这种理论只是停留在一般的分析水平上，回避了几乎所有的有待回答的重要问题。例如，尽管利菲弗里和皮纳德的基本程序是有效的，并且用同年龄的儿童做重复实验也获得了成功〔凯斯(Case)、1977年；利菲弗里和皮纳德，1974年〕，可是，当这种程序稍加变动后，它的效果就大大降低了〔福廷-西里奥特(Fortin-Theriault)，1977年〕。

那么，区分有效教学程序和无效教学程序的主要特点是什么呢？冲突—引导(Conflict-inducing)的经验在儿童心中引起了一个什么样的内部过程呢？以及如何最大限度地提高这个过程的效益？这又同样会引起一系列与成熟作用有关的问题。例如，利菲弗里和皮纳特(1972年)在他们最初尝试引导儿童进行阶段过渡时就注意到：许多年幼儿童似乎看不出估计数量的两种不同图式之间的内在冲突。按照第一次的观察方式，他们再用非常年幼的儿童做重复实验，继续试图复制第一次实验的结果，却没有成功。(福廷-西里奥特，1977年)。这是由于成熟带来的某种限制吗？如若这样，那么成熟是如何起作用的，在对教学进行设计时又应如何考虑儿童的成熟状态呢(利菲弗里和皮纳特，1974年)？

最后，我们提出一系列关于儿童平衡活动的作用问题。虽然利菲弗里和皮纳特的程序主要依靠设计一些练习来促进儿童自身内在冲突的解决，另一些包括更多教学措施的研究也获得了类似的重要结果〔布克(Bucher)和施奈德(Schneider)，1973年；格尔曼(Gelman)，1969年；齐默曼(Zimmerman)和罗森塔尔(Rosenthal)，1974年〕。那么，教学所引导的过程在多大程度上必定复制自发发展的过程？每个过程应包括些什么？有益的经验又是怎样影

响每个过程的？虽然，皮亚杰的理论在一般的描述水平上相当充分地阐明了认知发展的过程，但是，教学设计所要求的那种分析水平则要细致得多。

B.使教学适合学生现有的运算水平

一般地说，要使课程内容适合儿童现有的运算水平，也同样需要进行如下三步工作来促进运算的发展。唯一的不同之处是：第三步的目标不是要改变儿童现有的运算功能水平，而只是儿童在现有的功能水平上能学完每一个课程单元。然而，在实施时，每一步都会遇到严重的困难。

1.第一步：结构分析。我们首先要考虑的问题是分析不同教学任务和材料的基础结构。要记住，尽管皮亚杰没有明确阐述他的结构分析方法，但他隐约提到一组任务所要求的基础知识或能力，然后再根据一种抽象逻辑的观点来分析这种知识。可是，当人们的目的是要选择一种阶段——适合的方法或任务来教一个特定的概念时，所要传授的知识在正常情况下是要预先确定的。值得注意的是，皮亚杰的分析没有考虑到的正是这样一种变式。这种变式是指：从一种任务中占推断出的这种知识，能够从与这种任务相反的另一任务中很容易地推断出来。然而，问题不在于难以阐明或概括皮亚杰的分析方法，而在于，他的分析方法不适合首先要注意的目标。

2.第二步：个别评估。当我们进行个别评价时，也会遇到类似的困难。应当记住，由于皮亚杰的理论没有考虑到可能使所得结果发生混淆的“作业”或“任务”因素，这就给我们评估学生现有的运算功能水平带来了困难。当我们的目标是使课程材料适合儿童现有的功能水平时，该理论的这种缺陷就显得更为严重，因为儿童对这些任务变式的反应正是我们注意的焦点。

3.第三步：教学设计。最后，我们要考虑的是使儿童把现有的结构最有效地用于新内容中时所遇到的问题。由于皮亚杰的理论只注重于结构的获得过程，而忽视了结构的运用过程，因而从中就很少能得到直接的指导。的确，影响结构应用的任务因素同影响结构获得的任务因素原来是完全相同的。但是，由于皮亚杰的理论无论如何也没有阐明这些，因而最后的结果还是一样的。

总之，皮亚杰的理论如果是正确的，那么任何教学发展理论势必提出两个重要的问题：一是如何促进运算结构的发展；二是如何根据学生现有的运算水平进行文化上有价值的技能和概念的教学。一般程序似乎是清楚的，即：第一步分析所提出的教学内容本身的运算结构；第二步，确定学生当前活动所基于的运算水平；第三步，开发那些或则致力于促进这些结构进一步发展（如果是目标的话），或则致力于促进把这些结构应用到所关心的特定内容方面去（如果这是目标的话）的活动。虽然实现这些目标的一般程序是清楚的，但是要想把这些程序真正付诸实现时，每一步都会在理论上或技术上遇到巨大的困难。当旨在促进运算发展时，由于缺乏任务分析的一般图式和程序，给运算结构的识别带来困难。由于在皮亚杰学派的测量中可能存在着混乱的任务因素，给评估学生的现有功能水平带来困难；最后，由于对哪些成熟因素和经验因素能影响结构获得过程的图式缺乏详细阐明，而给最佳教学设计带来困难。企图选择一种适合发展的教学方法时，同样由于皮亚杰理论忽视了这种至关重要的"作业"或"任务"因素，而给教学方法的分析、学生对这些教学方法准备程度的评估、以及这些教学方法的最佳顺序，都将带来困难。

Ⅳ、新皮亚杰学派的认知发展理论

乍一看，前一部分所谈到的那些理论问题和实际问题似乎是很严重的，以致使皮亚杰理论完全不配作为建立切实可行的教学理论和教学工艺学的基础。尽管问题确实严重，但是应当看到，所有问题的产生都可以追溯到一个同样的根源，即：皮亚杰的理论主要注重的是结构而不是功能。正如弗雷沃和沃尔威尔（Wohlwill），1969年所指出的那样，皮亚杰首先关心的是对儿童在不同发展阶段上所共有的智力运算系统进行了逻辑描述，而没有对获得和运用这些运算系统的过程进行心理描述。假如皮亚杰的智力发展理论能够对获得和运用运算结构的过程作更详细的说明而加以补充，那么上述问题的严重性将会大大地减少。

首先，我们要考虑的是促进运算结构发展的问题。如有一个功能发展理论可用，那么我们就能更容易地鉴别出与各学科有关的智力运算。因为我们可以直接从儿童的实际作业中归纳出这些智力运算来，而较少地依赖于抽象的逻辑分析。此外，我们会更容易地评估儿童现有的功能水平。因为那些制约作业的因素会得到更好的理解，而所关心的评估类型与所观察到的作业会更加密切地关联起来；最后，我们还可以更容易地阐明促进运算发展的各种条件，因为这种阐明能够以对这些运算的最初获得过程的更加详细理解为根据。

接下来，我们要考虑使教材和教法适合学生现有的运算水平这个问题。如果有一个功能发展理论可用，那么：首先可以比较容易地按照学生发展的相应阶段来分析不同的授课方法和教材，因为教师所控制的任务变式类型将是这种理论详加论述的焦点；此外，还可以比较容易地评估那些决定学生从某种方法中受益的

能力的特点，因为决定儿童对任务因素做出反应的内部变量也能得到更好的了解；最后，可以比较容易地安排一系列活动以便把已有的结构最佳地运用到新的思想领域中，因为这种理论所要阐明的那些作业因素，不仅与结构的获得过程有最密切的关系，而且还与结构的应用过程有最密切的关系（假设这两个过程在实际上是不同的话）。

在这部分所谈的智力发展理论把皮亚杰对发展的解释看做是它的起点，同时还试图用阐明上述的那些教学问题所必需的功能理论加以确切的补充。这种功能理论最初是由帕斯考尔-莱昂(1969年)向日内瓦大学提交的一篇博士论文中提出的。因为这一理论的立足点是试图对影响皮亚杰所提出的具体结构的获得和运用的功能因素，作更加详细的论述。所以，这一理论的一些基本概念也只好在回溯到守恒任务的讨论以及对守恒任务的更加详细的功能分析之中来加以介绍。

A.液体物质的守恒：一种新皮亚杰学派的分析

在帕斯考尔-莱昂的理论阐述中，把儿童达到守恒之前所经历的那些阶段模拟成一系列复杂性和功力(Power)不断增加的心智策略(凯斯，1977a；帕斯考尔-莱昂1972年)。

1.策略1：单维扫描。处在皮亚杰所说的第一发展阶段(感觉运算阶段)的儿童，当要他们说明反应的理由时，他们主要着眼于两个液柱高度的差别。他们的眼动照片也表明：他们几乎完全沿垂直方向扫描这两个量杯〔奥布赖恩(o'Bryan)和博斯马(Boersma，1971年)〕。因此，他们做出反应所经历的心理步骤可归结如下：

(1)扫描量杯A中水的垂直维度。

(2)扫描量杯B中水的垂直维度（注意B杯中的水面总是超过A杯的水面）。

（3）确认B杯盛了较高水柱。

（4）得出结论说，B杯比A杯盛了更多的水。

在上述的每一步中，可以假定某些显著的图式是由知觉场直接激活的。另外，还可以假定儿童必须主动激活若干图式：或者是激活实现转换的运算图式；或者是不再由知觉场激活形象图式而是一定让这些形象图式仍然储存着。所有这些图式都逐条列入表3.1中。如表所见，必须照应的图式的最大量（不是由知觉场直接激起的那些）只是一个——这就是实现对高度进行比较的那个运算图式。如果高度的差别不明显，那么，图式的最大量就有两个，因为被试必须先计算出一个高度并把这个值储存起来，接着再计算出第二个高度并与第一个高度值进行比较，从而形成差别。

表3.1　　单维扫描策略的详细模型

步骤或运算	具体的图式	符　号
1.扫描量杯A中水的垂直维度	（1）扫描物体垂直维度的长度并贮存其最高点的运算图式	ψ 垂直扫描
	（2）b表示当眼睛沿着量杯A的侧面扫描时得到视觉输入的形象图式	φ A：侧面
2.扫描量杯B中水的垂直维度。注意量杯B中的水面总是超过量杯A的水面	（1）扫描物体垂直维度长度的运算图式。并注意它是否超过（仍是视觉的）先前的那个量杯A的水面的最高点	ψ 垂直扫描
	（2）b表示来自量杯B的视觉输入的形象图式注意到如果高度的视觉差别不明显，那么在这一步必须让另外一个表示量杯A的高度的图式储存着	φ B：侧面

3.结论是量杯B中的水比量杯A中的水面高	(1) 把看出不垂直投影上的差别做为输入并得出哪杯水面高的结论的运算图式	ψ 高度
	(2) b 表示看到的在第二步注意到的在垂直投影上的差别的形象图式	φ 在垂直长度上的视觉差别
4.结论是量杯B里的水比量杯A里的水多	(1) 表示物体越高容量越大这一规则的运算图式	ψ 高一些＝更多一些
	(2) b 表示在第三步中所产生知觉的形象图式，即量杯B比量杯A高。	φ B＞A(高度)

a、用符号ψ表示一种运算图式，用符号φ表示另一种形象图式。

b、在帕斯克尔–莱昂的体系中，假设视觉上被激活的图式并不以心理能量的激活作用为必要条件。因此，这些视觉图式就不用从心理需求的估计方面去加以考虑。

2.**策略2：双维扫描**。当让儿童说出他们的反应的理由时，处于皮亚杰第二个发展阶段的儿童有时说到高度差别，有时则说到宽度差别，这就有赖于哪些特点更鲜明。如果两方面的差别都不很明显，那么他们往往要花很长时间决定哪个差别更明显些。眼动照片表明他们对两个量杯进行了比较全面的扫描，而且并不局限于垂直方向，因此，他们完成回答所经历的心理步骤的顺序可以描述如下：

(1)扫描量杯A中水的垂直维度。

(2)扫描量杯B中水的垂直维度(注意B杯中水总是超过A杯的水面)。

(3)确认量杯B容纳较高水柱。

(4)如果垂直维度的差别大，那么结论是B杯的水比A杯的水多。否则，将进行第五步。

(5)扫描量杯B中水的水平维度。

(6)扫描量杯A中水的水平维度。

(7)注意量杯A容纳了一个较宽的水柱。

(8)如果水平维度的差别大，那么结论是量杯A的水比量杯B的水多。否则，再回到第一步，确立一个低值的"大的"标准。

在上述过程中每一步所必须激活的图式都列在表3.2中。如表所见，必须激活的图式的最大量只有两个——同表3.1相比较就新增加了一个，这由于如下的事实：即关于两个物体哪个较宽这个问题并没有直接的视觉线索可供参照，如果在宽度和高度的差别之间作出直接的比较，那么，必须激活的图式最大量就是三个，因为在估计和比较宽度时，高度的比较结果将不得不贮存起来。

3.策略3：根据状态进行推理。如前所述，那些做出守恒反应的儿童，在说明他们的答案的理由时，几乎总是说到液体最初的相等和转换的性质。眼动照片也表明，他们对转换后陈列的扫描常常是十分粗略的(奥布赖恩和博斯马，1971年)。因此，他们作出回答所经历的一系列心理步骤可以描述如下(帕斯考尔-莱昂，1972)：

(1)回忆两个量杯初始的相对数量，贮存。

(2)回忆转换的性质。

(3)如果这种转换只是倒入，那么最终的相对数量与最初的相对数量相同。

上述每一步必须被激活的图式都列在表3.3中。如表所见，在任何一步中必须激活的图式的最大量是三个——一个表示初始相等；一个表示倒入这一事实；还有一个表示"倒入并不改变数量"这一规则。

表3.2　　双维扫描策略的详细模型

步骤和运算[a]	具体的图式	符　号[b]
5．扫描量杯B中水的水平维度	(1) 扫描一个物体的水平范围，并贮存其数量的运算图式(也许作为一种表象，或作为一种扫描时间的记录)	ψ 水平扫描
	(2)[c]当眼沿着量杯B的底部扫描时，表示所得视觉输入的形象图式	ϕ B杯底部
6．扫描量杯A中水的维度并把它同量杯B作比较(注意：因为和高度一样，对于相对程度并没有视觉线索，量杯B长度不得不暂时贮存着)	(1) 表示量杯B宽度的形象图式(在第一步形成的)	ϕ B杯的水平量度
	(2) 扫描物体的水平范围并把它的数值与前面贮存的值进行比较的运算图式	ψ水平扫描
	(3)[c]表示来自量杯B底部的视觉输入的形象图式	ϕ A杯：底部
7．注意量杯A是比较宽的。	(1) 把两个水平距离当作输入并把适当的一个称为宽度的运算图式	ψ 宽度
	(2) 表示A杯的水平数值的形象图式(在第五步形成的)	ϕ A杯的水平量度
	(3) [c]表示B杯的水平数值的形象图式(在第六步形成的)	ϕ B杯的水平量度
8．结论是量杯A中的水较多。(除非在第7步难以看出差别，如果是那样，回到表3.1的第1步。	(1) 表示物体越宽容量越大这一规则的运算图式。	ψ 宽的＝多的
	(2) 表示在前一步中形成的A杯比B杯宽这个结论的形象图式	ϕ A＞B宽度

a．1—4步的运算见表3.1

b．ψ＝运算图式　　ϕ＝形象图式

c．由知觉的输入激活。

表3.3　　根据初始状态进行推理策略的详细模型

步骤和运算	具体的图式	符　号ᵃ
1. 回忆两个量杯最初的相对数量	(1) 表示在开始时就判断出A杯中的水量等于B杯中的数量这一事实的形象图式。	φ A=B
2. 回忆转换的性质	(1) 表示初始相等的形象图式(见第一步)	φ A=B
	(2) 表示水从B杯倒入B'杯并没有增加或减少这一事实的形象图式	φ B倒入B' ⟶
3. 推导出关于现在相对数量的结论	(1) 表示如果初始状态相等，没有增加或减少，那么最终状况也相等这一规则的运算图式。	ψ 倒入
	(2) 在第二步中产生的形象图式	φ B倒入B' ⟶
	(3) 在第一步中产生的形象图式	φ A=B'

a.ψ=运算图式　　φ=形象图式

现在我们来考虑阶段过渡的问题。按照皮亚杰的看法，我们可以假设阶段过渡有两个一般的机制。第一个机制是一个比较缓慢而带有或然性的机制，凭借这种机制儿童逐步觉察到了在第一个策略中没有考虑到的问题的维度，并逐步探索出处理这种维度的某种新策略；或者，由于儿童运用了第一个策略，他可能体验到某种认知冲突，或者因为他没有达到某一外在的目标，或者是因为他遇到了一种内在的矛盾——而这种冲突的体验能使儿童去积极地探索某种新的维度，并把这种维度考虑在内去拟定出一种新的策略。从教学的观点来看，这后一种过程是更有意义的，因为它更加符合可由直接干预模拟的那种顺序。

因此，我们来考虑一下帕斯考尔-莱昂的理论体系中这种过渡

是如何模型化的。尽管因运用第一种策略或第二种策略体验到冲突的情况有多种多样，但是，所看到的最简单的情况可能是处在第二个发展阶段的被试的那种情况，如果在某些情况下他使用高度作为估计的基础得出一种答案，而如果他使用宽度作为估计的基础就得出另一种答案。通过下面一系列心理步骤，可以得到上述的认识(凯斯，1977年b，1976年)：

(1)注意量杯B中的液柱比量杯A中的液柱高。

(2)根据这个标准，结论是B杯盛了更多的水。贮存起来。

(3)注意量杯B比量杯A窄。

(4)根据这个标准，量杯B盛了较少的水。贮存起来。

(5)看出两个所贮存的结论是互助冲突的。

上述每一步中必须激活的图式列在表3.4中。如表所见，在每一步中必须激活的图式的最大量是三个：其中两个是为了得出关于宽度的结论的；另一个是为了贮存关于高度的结论的。

表3.4　运用双维扫描策略时发现矛盾的详细模型

步骤和运算	具体的图式	符　号。
1．注意B 杯比A 杯的液柱高	(1) 把看到的两个物体之间垂直投影的差别作为输入和把哪一个较高的结论作为输出的运算图式	ψ 高度
	(2)[b]表示所看到的两个液柱之间垂直移位的差别的形象图式	φ 垂直移位
2．结论是在B 杯中的水多	(1) 表示越高的东西容量越大这一规则的运算图式	φ B>A (高度)
	(2)[b]表示在第一步中产生的事实，即B 杯比A 杯高的形象图式	ψ 高=多
3．注意B杯比A杯细	(1)[c]在第二步中产生的形象图式。这表示B杯比A 杯容量多一结论。	φ B>A (高度)

	(2) 把所看到的水平投影差别作为输入和把哪一个校宽的结论作为输出的运算图式(见表2的详细资料)	ψ 宽度
	(3) 表示B杯和A杯水平投影的形象图式(关于图式产生的详细资料见表2)	φ 水平位移
4. 结论是B杯中的水少	(1)[c] 在第二步中产生的形象图式(B>A)	φ B>A
	(2) 表示宽的东西容量多这一规则的运算图式	ψ 量(宽度)
	(3)[b] 表示B杯比A杯细这个事实的形象图式	φ B<A (宽度)
5. 注意到两个所贮存的结论是互相冲突的	(1)[b] 在第二步中产生的形象图式(B>A)	φ B>A
	(2) 在第四步中产生的形象图式B<A	φ B<A

a. ψ=运算图式　　φ=形象图式

b. 由知觉输入而激活的。

c. 为以后用而激活的图式(贮存)。

儿童一旦对他目前的策略感到不满，他就可能开始探索一种更合适的策略。儿童根据初始状态发现推理策略可能经历的一系列步子如下(帕斯考尔-莱昂，1972年)：

(1)记住最初的两个量杯(A和B)中开始装着等量的水。

(2)记住把量杯B的水倒入高而细的量杯(B')中，既没有增加水，也没有减少水。

(3)根据过去倒水的知识中引出这样的结论：量杯B的水在倒入B'杯的过程中，数量并没有发生变化。

(4)将B'仍等于B和开始时A就等于B这两个事实结合在一起考虑，所得的结论是A杯中的水量等于B'杯中的水量。

上述每一步必须激活的图式列入表3.5中。由表可见，必须激活的图式的最大量还是三个。

表3.5　获得守恒的详细模型

步骤或运算	具体图式	符　号[a]
1．记住A杯的水量＝B杯的水量	(1)[d]表示实验开始时A杯的水量＝B杯的水量这一事实的形象图式〔有关对开始引出这个结论所须运算的说明见Pascual-Leone(1972)〕	φ A＝B
2．记住B′杯中的水是由B杯倒入的	(1)[b]表示A杯的水量＝B杯的水量的形象图式见上	φ A＝B
	(2)[b]表示所观察到的变化，即B杯中的水倒入B′杯中，液体并没有增加或者减少的形象图式。	φ B倒B′ ⟶
3．推断B杯的水量＝B′杯的水量	(1)[b]表示A杯的水量＝B杯的水量的形象图式(见第二步)	φ A＝B
	(2) 适用于各种变化并得出水量有无变化这一结论的运算图式，这种图式就相当这一规则：如果没有增加或减少什么东西，那么数量保持相等。	ψ 倒＝相等
	(3) 表示变化液体方式的形象图式	φ B倒B′ ⟶
4．结论是因为A杯的水量＝B杯的水量，而B杯的水量＝B′杯的水量所以A杯的水量＝B′杯的水量	(1)[c]表示如果A＝B，并B＝C，那么A＝C这一传递规则的运算图式	ψ 变化
	(2) 表示从上一步得出B杯水量没变，B杯水量＝B′杯水量这一结论的形象图式	φ B＝B′
	(3) 表示开始状态A＝B这种记忆的图形图式	φ A＝B

注：改编自J.帕斯考尔—莱昂：《一种结构算子理论，新皮亚杰学派的一个守恒模型，和横向发展阶段问题》，未发表的手稿，约克大学，1972。

a.ψ＝运算图式　φ＝形象图式

b.供下一步用的激活的图式(贮存)。

c.也可以不要运算图式，所以得出这一答案是因为当A＝B′与B＝B′和A＝B成立时，要比 A≠B′，表示了一个更加简单和更加一致的概念领域〔关于这种可能性的说明见帕斯考尔-莱昂(1972年，1974b)〕。

总之，上述所有模型，不仅可以看作是对儿童在获得守恒过程中所经历阶段的程序的一个说明，而且也可以看作是当儿童遇到一种适宜的环境条件时从一个阶段过渡到下一个阶段所依赖的那种过程的一个说明的。这些模型的细节是带有推测性的，还没有加以检验。而从本文的观点来看，要注意的重点，不是好歹提出的某一套心理步子或图式，而要注意的重点是一旦采用了这种模拟儿童发展的方法而显然变得适当的那组作业因素。如果这种方法是有效的，哪怕是大致有效，那么我们希望下面将要阐明的一系列因素能影响儿童获得守恒概念的速度。

第一种因素是：儿童同时照应几个图式这种能力的发展速度。不管所假定的具体图式集合如何，有一点是很清楚的，即执行第一个心智策略比执行后继的一些心智策略必须同时激活的图式要少。因此，儿童所能激活图式的最大量与他们对守恒自己曲解之间应有着一种关系（帕斯考尔-莱昂，1969年）。最近的实验研究为这种主张提供了论据。实验结果表明，儿童到了完成守恒的年龄一般才能同时激活三个图式，(凯斯，1972年b；帕斯考尔-莱昂，1970年)。实验还表明，甚至在过了完成守恒的年龄之后，这种能力测验同守恒测验也有显著相关。

影响儿童成败的第二种因素是易受知觉定势的干扰，还是能抵制知觉定势的影响。探讨守恒问题的前两个策略都包括对事物的鲜明的知觉特点的评价。相比之下，最后一种策略要求儿童把他们的注意力从当前的视野中移开而很据以前的知识作出回答。因此，场依存性的认知风格〔威特金(Witkin)戴克(Dyk)，法特森

(Fatersen)古迪纳夫(Goodenough)和卡波(Karp), 1962;威特金,刘易斯,赫茨曼(Hertzman),麦克奥瓦(Machever),迈斯纳(Meissner)和韦普纳(Wopner) ,1954〕与守恒任务作业之间应该是相关的。在帕斯考尔-莱昂攻读博士学位的研究中,证实了这种假设(帕斯考尔-莱昂,1969)。他也看到了场依存性与守恒作业之间的关系显然很密切,一个有明显的场依存性的聪明成人也常常不能完成守恒任务。

影响儿童成败的第三种因素是,关于液体、量杯以及相对数量判断的已有经验。例如,如果儿童遇到许多种作用相反的高度数量和宽度数量进行比较的情境,那么他们比没有遇到这种情境更有可能经历表3.4中详细列出的一系列心理步骤。同样,如果儿童遇到这样的情境,即向他们指出双维策略固有的困难并要仿照更为复杂的策略,那么他们甚至更可能经历表3.4中详细列出的一系列步骤并采用表3.5所概括的策略。第一类情境的经验给儿童的发展提供了机会;第二类情境的经验给儿童的发展提供实际指导。最近的实验研究证明了这两种经验都是有益的。如果仅是向儿童提供大量不能预见结果的双维评价的情境,那么,儿童就能自发地形成一种守恒的理解——也就是说他们至少能同时照应三种图式(凯斯, 1977年a)。此外,如果主动向儿童指出双维策略的不适当性并指导他们掌握正确的策略,那么,他们就能更容易地,甚至在较低水平心理——功力(M—Power)上获得守恒(Case 1977,a)。

最后,影响儿童成败的第四种因素是他的情感素质。如果儿童被测验情境吓倒,那么向他们提出守恒问题时,他们就不可能将他们理解的情况作出真实的说明。而更重要的是,如果有引起儿童忧虑的数量判断情境,那么他们开始时就不可能很全面地去探索情境或者经历上述各表中详细列出的一系列心理步骤。尽管这后一种可能性似乎极小,但必须记住,激励他们经历表3.5详细列

出的一系列心理步骤的动机是在经历表3.4中详细列出的各步的过程中所体验到的那种认知不确定性。还要记住，不确定性的情境原就是一个由情感引起的情境，不同的儿童可作出不同的反应(凯甘，1971年)。

当我们对皮亚杰学派的另一些问题(如类包含的问题)，进行上述详细分析的时候，这四种因素也常常是决定儿童获得成败与否的基本因素。在此分析的基础上，导致帕斯考尔-莱昂去建立关于儿童获得和运用皮亚杰所说的那些认识结构所凭借的过程的一般理论。尽管在已经提到的有关守恒问题的分析中，这个理论的主要原则还是含蓄的，但是，在进一步研究之前是值得花时间把这些原则弄明确的。

最近，已就这一理论做了许多增补和修改(凯斯，1978年b；帕斯考尔-莱昂，1976年a)。可是，为了满足当前的需要，只总结了最稳固确立的并与教育有最直接关联的那些方面。此外，描述该理论所用的语言将尽可能接近于本卷书其余章节所用的语言。有关这一理论的更完整的论述，见帕斯考尔-莱昂(1970年，1972年，1976年b)或安蒙(Ammon1977年)的著作。有关对这一理论所提出的补充和修改的论述，见凯斯的著作(1978年a)。

B.知识的基本单位

正如在守恒分析中所阐明的那样，人类的全部知识，无论是事实性的还是程序性的，它们都可以假定是用称谓图式这种实体贮存于心理系统中的，如同计算机模拟研究者所假定的知识单位一样(纽厄尔和塞蒙，1972年)，我们可以假设图式是由两种成分组成的，即：图式应用的初始条件集合(图式的释放成分)和图式产生成的后继条件集合(图式的效应成分)。例如，我们认为大多数4岁儿童在完成守恒任务的开始阶段就表现出这种能力，比较两个量

杯并得出结论说这两个杯盛的水相等。在这种理论体系中，我们假定存在这种能力，因为儿童在他们的认知指令系统中已经具备了相等这一概念的图式。这一图式的效应成分是指儿童在其特定的发展阶段赋于相等这一概念的特定意义；而图式的释放成分是指对这种意义进行合理赋值的可能条件集合。用上述理论语言，我们要么可以说儿童关于初始相等的结论是由刺激情境“释放”的，要么可以说儿童把相等概念的理解运用到刺激情境中并符合要求。

也正如在守恒分析中所论述的那样，图式可有各种不同的功能。凡具有表征功能的图式——表征事实、状态或意义的图式——都可以称作形象图式；凡具有转换功能的图式——对一组形象图式进行运算而产生一个或一组新的形象图式——都可以称作运算图式；凡具有控制功能的图式——表示一个被试为了把一种形象图式状态变成另一种形象图式状态而要执行一系列运算图式的图式——都可以称作执行图式。

我们并没有假设这三种图式之间的差别有其结构基础。在不同时期，根据所提出的问题，同一结构单位可以具有这三种功能中的任何一种。而从分析的目的出发，有一个规定某一图式的功能的术语是很方便的，上述所用术语的优点是同现代认知心理学中所用的那种加工特点相一致的，同时也与它所派生的皮亚杰学派的术语相符合的(帕斯考尔-莱昂1977年b，皮亚杰，1969年)。

C.智力发展的内容

在帕斯考尔-莱昂的理论体系中，可以把皮亚杰所说的结构顺序看作是功力和复杂性逐步增加的一系列执行图式。正如前面的分析中所指出的那样，关于守恒的执行图式一定与数量的估计有关。

开始时，被试是在单维的基础上进行数量估计，一般只看高度。以后，当他们发觉这样做不总是行得通的时候，他们就寻求改进评价的依据(克拉尔和沃利斯，1973年；皮亚杰，1957年)。他们先是交替使用宽度和高度，然后是高度和宽度联合使用。虽然双维执行办法是更为有效的，但是，在某些情境中并不合适，尤其是当高度和宽度这两个矢量在数量上接近而作用的方向相反的时候，进行补偿比较就更不合适了。当被试感到后一种比较有许多矛盾和困难的时候(可能经历如表3.4中所描述的过程)，他们就再一次寻求一个更好的定量基础。最后，无论是在他们的日常经验中，还是在守恒测验的情境下，他们都要经历表3.5中所列的那些心理步子，并且发现可用大小相同的容器作为测量标准来确定相对数量。尽管上面所谈到的那些专门的执行图式是很不一样的，但是可以看出有一个类似的事件序列说明了在其它各种皮亚杰问题中所观察到的变化。〔诺埃尔廷(Noelting)，1975年；西格勒(Siegler)，1976年〕。只要把智力发展的内容看作是功力和复杂性逐步增加的执行图式的一种序列时，那么就有可能把表3.1至3.5中所描述的各种模型作为一种原型，并提供一个有关平衡过程的一般特点。

D.平衡过程

在新的理论体系中，我们可以认为下列的一般设定事实代表了皮亚杰称之为平衡的那种自发冲突解决的特征：

1.当在某种目的一指向活动中，两个实际上不相容的图式被一并激活起来的任何时候(例如：图式1—A杯有更多的水；图式2—B′杯有更多的水)，有机体就体验到认知冲突。

2.当有机体体验到这种冲突时，就暂时放弃它当前的执行图式，并开始寻找可以帮助它解决冲突的任何其他信息。用计算机模

拟的语言来说，它激活了一个搜索性执行图式(HES)。这个图式指导对问题情境、问题已知条件和所有存贮的任何有关信息的探索(最初的搜索性执行图式的发展大概发生于感觉运动阶段，儿童到了一岁半当他运用一种策略无效而受挫折时，我们会看到他就积极地去探索另一种策略)。

3.有关信息的搜索是按零散的“心理步子”进行的，其中每一步都花了一定数量的实际时间。〔这种时间的最大单位相当于所谓的“心理时间”(斯特劳德(Stroud)，1955年)〕。

4.在每一心理步子中，或者是(a)一个形象图式或图式集合被检索，或者是(b)把一个运算图式应用到某个已激活的形象图式上并产生一个新的图式。

5.任何图式的激活或检索都需要运用“心理能量”(mental energy)。除非有即时的输入的某一突出方面的直接支持。因为在任何一个时刻可供运用的心理能量的数量是有限的，所以在任何一个心理步子所能激活的图式数目也是有限的。我们把这种数目称为有机体的心理一功力(M—Power)〔虽然这种结构与短时记忆(STM)十分相似，但我们必须指出它们之间存在着几个重要的区别。从理论上说，短时记忆是根据外部反应而不是根据内部图式界定的。因此，由感觉作用(如听觉)激活的图式应该用短时记忆的估计数来计算，而不应该用心理一功力的估计数来计算。从经验上说，短时记忆测验只要求被试感知和再现若干单位，而心理一功力的测验则要求以某种方式来转换诸单位，因此，这种区别类似于短时记忆与操作记忆之间的区别〔波斯纳(Posner)，1970〕。

6.如果冲突没有消除，而其它条件又相等，有机体就乐于作出同当前激活的最大量的图式相一致的反应。

7.有机体对所体验的冲突进行几次尝试而得到解决之后，如此的心理步子序列就作为一种策略被固定下来了，并且被试能根

据以后的尝试去产生新的答案，而不需要任何搜索性执行图式的帮助。简言之，有机体形成了新的执行图式(ESZ)。对这种图式来说，最初的问题情境是一种释放器(releaser)(如：在高度上A＞B，在宽度上B＞A)，而最后的解决构成了效应器(effector)(如：检查转换前的状态，检查转换的类型；如果一开始相等，并且没有增加或减少，那么其结论是仍然相等)。〔有关这种学习发生条件的正式的详细论述，见帕斯考尔-莱昂，帕金森(Parkinson)和普洛斯(Pulos)1974年〕。

E. 影响智力发展速度的因素

根据这种平衡的一般特征，可以看到有四种性质截然不同的因素将影响被试智力自然发展的速度。这些因素是：被试的心理—功力的大小、他的场独立性程度、他所具有的有关内容的经验以及情感因素。

1. 心理一功力的大小。第一个因素是被试心理一功力的发展速度。在守恒任务中，我们应当记住，被试看出冲突并找出解决冲突的答案所需要的心理一功力最小量是3。在皮亚杰学派的另一些任务中，心理一功力最小量有时高达5或5以上，有时则低到2。在这种大量分析的基础上，同时考查了正常通过所谈任务的年龄，帕斯克尔-莱昂假设心理一功力是按表3.6所表明的比例随年龄线性增长的。他还假设心理一功力的增长主要依赖于成熟的因素或者一般的经验，也就是说，它不受特殊经验的影响。

2. 场独立性。可影响被试智力发展速度的第二个一般因素是被试的场独立性的程度。在守恒任务中，应当记住两个量杯看起来好象所装的液体数量不等，这可能是因为液柱弯月面特点如此突出，而液体的表面面积的特点不突出，就高度差和表面面积差来说，甚至连成人也倾向于过高估计高度差的重要性。因此，如

表3.6　　不同年龄心理—功力值的假设模型

年龄(岁)	皮亚杰学派的亚阶段	M的值[a]
3—4	前运算早期(直觉思维)	e+1[b]
5—6	前运算晚期	e+2
7—8	具体运算早期	e+3
9—10	具体运算中期	e+4
11—12	具体运算晚期 形式运算早期	e+5

a. 帕斯考尔-莱昂(1970年)提出在15或16岁时心理—功力可以增长到e+7，但还没有什么相应的文献证明这个年龄范围心理—功力的增长情况。

b. 符号e表示执行所占的间隔，假定它是恒量。

果他们没有亲眼看到最初的相等和变化的情形，那么，他们就可能倾向于作出这样的判断：含有较高液柱的量杯能装着更多的饮料(凯斯，1976年)。由此可见，深受知觉场结构影响的被试，对于不确定数量的无论哪一种知觉基础的执行图式所固有的困难都不敏感。因此，他就要花较长时间去寻找一种能建构一个更合适的执行图式的方法。实际上，其它许多皮亚杰学派的任务都同样具有这种特点，而这种形式特点就成了测验运算发展的一种常用标准。在皮亚杰的体系中，知觉场可以引起一个不正确的反应，这确是用一种可靠的方法认识到：正确反应不是来自对某一个突如其来的刺激的自动反应(皮亚杰和英海尔德，1941年)。

已知某些皮亚杰的任务（不是全部的），会使被试产生错误的场效应，因此，我们可以假定，特别易于受这种效应影响的被试在显示这种效应的那些内容方面发展精细的执行图式要慢于正常水平。但是，在其它内容方而发展这种图式却并不比正常水平慢(帕斯考尔-莱昂，1969年)。

3.有关内容的经验。影响被试智力发展速度的第三个因素是他们在任一特定内容方面所经历的经验的数量和质量。在守恒任务中，须记住有两种经验是至关重要的：一种经验是为儿童提供了放弃知觉策略而使用比较理性的策略的机会；另一种经验是在完成守恒任务的过程中为儿童提供直接指导。对大多数其它皮亚杰任务也同样如此，既可通过为儿童提供建构更合适策略的机会的经验，也可以通过为他们提供行动指导的经验，来使作业易于完成。

关于经验在影响儿童认知发展中起着重要作用的这一主张当然不是新的。皮亚杰及其追随者在某一个时期也曾强调了能为儿童提供学习机会这种经验的重要性。他们曾假设这种经验的效力主要是通过为儿童提供机会练习每个图式并发现这些图式的局限性，或通过诱发两个显然矛盾的图式之间的平衡而实现的（英海尔德等人1974年；皮亚杰，1964年，1971年；塞劳斯和兰格，1970年）。同样，新行为主义的研究者们在某一个时期内也曾强调能为学习过程提供直接指导的这种经验的重要性。他们也曾假设这种经验的效力是通过引起儿童注意一个更适当的反应所须依据的提示上，(格尔曼，1969年)，或是通过提供一个这些提示必须如何处理的清楚模型而实现的(贝林，1965年；齐默曼和罗森塔尔，1974年)。

新主张的不同之处是强调这两种经验的潜在重要性。第一种经验在于降低了被试使用当前执行策略定势的重要性；第二种经验的重要意义在于促进发现这种执行策略所存在问题的过程并建构一个更合适的策略。因此，我们认为，第一种经验是智力自然发展的主要原因——也就是说，在这种发展中儿童的学习是受他自己的搜索性指令系统控制的，并且经受着结果中全部难以预测的变化。相对而言，我们认为第二种经验则是通过教学促进发展的主

要原因——也就是说，在这种发展中儿童的心理步子是由已经知道什么策略更合适的人进行指导、简化和监督的。

从新的理论观点出发，两种传统的思维学派之间的争论实际上是一种互相的误解。发展学家已探求自然发展的机制，并得到经验只具有适度作用的资料(英海尔德等人，1974年)。相比之下，学习理论家则研究了通过教学引起发展的机制并得到经验有巨大作用的资料(布克和施奈德1973，格尔曼，1969)。

4.**情感**。影响被试智力发展速度的第四个也是最后的一个因素是情感。在某些情境中，善于探测人类的生存和幸福是特别重要的，我们对这些情境特别喜爱，因为这些情境能立即强烈地激活有机体所具有的用来处理它们的任何执行图式。一个从情感上激活的图式可完全优先取得任何正在控制行为的其它执行图式(塞蒙，1967年)，或者这两种图式以复杂的方式互相作用。①

乍看起来，相对地说，似乎情感因素在智力发展的过程中作用并不大。但是，正如在守恒任务的分析中所指出的那样，实际情况并非如此。例如，完全可能是儿童对某些内容方面产生一种情感反应，这种反应就使他在心里把这一内容方面与其它内容方面分离开。此时，儿童或者对此感到特别兴奋，或者感到特别厌烦，或者忧心忡忡。在此情况下，儿童在那一内容方面的发展速度以及他所能达到的最终水平会因此而加速或者减慢。

但是，更重要的是，儿童对可能获得执行图式的情境本身体验到一种情感反应。正如皮亚杰常常指出的，也象新皮亚杰学派在阐明平衡过程时所强调的那样，一种能对智力自然发展起促进作用的情境是认知不确定或冲突的情境。如果没有经过冲突就获得新的执行策略，如果有足够的心理——功力去组合一个克服冲

① 关于在低等动物中这种复杂的相互作用的描述，见Hess(1962)。

突的执行策略，如果他们对一个减少冲突执行策略的成分不经过漫长的探索，那么他们所体验的冲突和不确定性会转瞬即逝或毫无意义。而如果把儿童置身于冲突的情境中：(1)他们没有足够的心理一功力去组合一个更有效的执行策略；或者，(2)他们探索适当的成分是很困难的；那么，此时他们体验到的冲突和不确定性就可以变得持久而具有相当重要的意义。因为不确定性和冲突构成了人们生来就对之表现出情感反应的情境(凯甘，1971)，所以，儿童一般智力发展的速度和最终水平明显地受他对不确定性的情感反应的性质所影响的。由于文化因素、家庭因素或个人习性因素的影响，一种情况是儿童将不顾一切地去学会避免这种不确定性和矛盾，或者使得儿童在寻求冲突的解决时变得胆怯和一昧等待成人提供答案。另一种情况是，也由于这些因素影响的结果，儿童可以学会主动地去发现不确定性和冲突，他可以相信他有能力自己去解决这些冲突，为此，他独立大胆地发展各种策略。在前一种情况下，儿童的发展速度是比较慢的，而在后一种情况下，儿童的发展速度就比较快。

F.皮亚杰理论与新皮亚杰学派理论之间的关系

从总体上看，无论是新皮亚杰学派的体系还是传统的皮亚杰体系，有关智力发展过程的特点都是完全相同的。然而，正如读者无疑会注意到的那样，在界定主要结构的方式精确程度上，它们之间却有着许多重要的区别。

1.运算结构。两派理论都把运算结构作为一个基本概念，把它界定为反思已有知识、并通过对平衡过程的控制而支配新知识的获得的图式集合。在两派理论中，皮亚杰的理论把运算结构作为逻辑能力的模式，其数学性质可借助于符号逻辑作出最好的说明。而新皮亚杰学派的理论却把运算结构作为执行图式集合的模

式，其释放线索和有效动作序列，可以直接从儿童的作业中推断出来，并能用标准英语加以详细说明。

一组执行图式的概念和一个逻辑结构的概念是完全一致的〔(塞勒里尔(Cellerier)1972年)〕。但是，由于一个执行图式的内容比一个逻辑结构的内容能更容易更具体地进行详细说明，因此这样在概念上的转换，就可使儿童思维模式形成的过程变得更详细、更精确并与可能观察到行为联系更密切。如同可以把运算结构等同于执行图式的概念那样(塞蒙，1962)，由于执行图式概念起源于计算机模拟的研究〔纽厄尔肖(Shaw)和塞蒙，1958〕，因此，把皮亚杰所认定的发展阶段作为执行图式序列的模式，就可在发展心理学的研究与人类信息加工的研究之间架起一座概念的桥梁(凯斯，1970，克拉尔和沃利斯，1973；西格勒1976)。

2.平衡。皮亚杰理论中把平衡作为一个基本概念，它指的是一个由认知冲突所引起的图式改组的过程。新皮亚杰学派的理论则把平衡作为一种特定类型的问题解决的模式。这种问题解决过程的独特性，在于它是由有机体当前执行的指令系统所产生的不一致性或冲突所引起的；还在于它导致那个指令系统发生了变化。不过，就其基本过程来说，它同其他一切问题解决过程一样，都遵循着同一套一般法则。再则，尽管在概念上与皮亚杰的思想是完全一致的，但新皮亚杰学派理论所赋与平衡的新特点，却使得对过程的分析更加丝丝入扣——这种分析所用的时间框架十分短；分析与儿童在特定情境中的行为更加密切。另外，给平衡赋予新特点也就把古典的发展理论和现时的信息加工理论紧密地联结起来了。

3.成熟和一般经验。在传统的皮亚杰学派理论体系中，并没有详细论述成熟和(或)一般经验影响智力发展的过程。相反，在新理论中，假设这些因素是通过改变有机体的中心场或心理——

功力的大小而起作用的。另外，新的定义比传统的皮亚杰体系能够对所谈到的运算结构作更准确、更具有运算性的说明。结果是在特定的背景下可预见成熟限度的所在。另外，所提结构性质再一次在传统的发展理论与现行的人类信息加工理论之间架起了一座概念的桥梁[布罗德本特(Broadbent),1973；卡尼曼(Kahneman)1973；米勒，1956；帕斯考尔—莱昂，1970]。

但是，与运算结构可等同于执行图式的情况不一样，或者也与平衡过程可以作为另一系列指导问题解决的心理步子的情况不一样。把成熟的限度等同于心理—功力的限度就使皮亚杰理论的基本结构发生了两个重要变化：第一，这一假设把所提出的皮亚杰称为中心场的大小和运算结构复杂性之间的关系颠倒了。对皮亚杰来说，儿童中心场的增长是其运算结构复杂性增加的一种结果，而不是一种原因（皮亚杰，1920年，1928年)；第二，对心理—功力的强调改变了假定会限制获得新结构速度的因素。对皮亚杰来说，这类限制因素是指平衡固有的缓慢性以及运算改组任务的大小(皮亚杰，1970年)。而对帕斯考尔—莱昂来说，唯一的限制因素是中心场大小的限度(心理—功力)。正如在下面两部分将要进一步阐明的那样，这种重点转移的结果将对教学在影响智力发展过程中所起作用的评价发生很大改变。

4.知觉的作用。在皮亚杰的理论中，知觉的作用是作为认知任务变化的一个可能原因提出来的。但是，它并没有清楚地阐明知觉起作用的机制。相比之下，在新皮亚杰理论中，知觉作用则是从格式塔心理学家研究过的那些刺激场的感觉组织特性派生出来的基本概念。假定这些因素是通过影响执行图式的选择而起作用的，并且在某些情境下，这些因素由此能影响对一种任务的心理—需要。进一步还假定：对这些因素感受性的个别差异是稳定的，并相当于Witkin及其合作者曾经研究过的认知风格，另外皮

亚杰学说的这些发展将允许对特定任务的作业作出传统皮亚杰理论不可能作出的预测〔凯斯和格洛伯森(Globerson), 1974帕斯考尔一莱昂,1969;斯卡达马利亚(Scardmalia),1977C〕。

5.外部诱导的学习(VS) 与自发的学习。在新的理论中谈到了通过社会经验从外部诱导的学习和正常发展过程中自发产生的学习这两者之间的区别。然而，这两种类型学习之间的异同乃得更加清楚、更加着重从运算方向来加以详细阐明：结果呢，一种特殊的经验就可以归为维持日常经验发展需要的一类(即维持运用一种搜索执行结构的相同需要和要求相同的心理一功力)，或者可以归为由外部诱导造成缩减这些发展需要的一类(即促使搜索执行结构的寻找工作容易进行和缩减心理一需要)。

6.情感。在两种理论中，情感都是作为一种潜在的重要因素而被提出来的。但是，在新的理论中，假定这种建构起作用的方式，促使情感的影响在特定条件下会(至少有可能)更易于识别出来。因此，就更加容易与心理学的其他领域的研究联系起来〔如，艾贝尔森(Abelson)，1973年；赫斯 (Hess)，1962年；斯坦纳(Steiner)，1974年〕。

总之，由帕斯考尔一莱昂创立的理论对传统的皮亚杰学说作了许多精巧的改造，从而有可能更加容易和精确地识别出关于构成皮亚杰学派任务作业的运算结构，并有可能更加详细地阐明影响这些结构获得的种种因素，如成熟、知觉、经验和情感等。经过精巧改造后建立的新理论，其中大部分仍保持了传统皮亚杰理论的基本结构。但是，在把成熟的限度等同于心理一功力的限度这一方面，新理论提出了一条重要的不同意见。

V. 关于新理论的研究

尽管新的理论还没有开展大量的研究，但是已做的研究就已经为它的基本假设提供了重要的支持。在这一节中，我论述这方面的研究，而在随后的第Ⅵ和第Ⅶ这两节中，我将论述新皮亚杰学说对教学的意义。

A.心理一功力随年龄而发展

由于我们把心理一功力是定为在无知觉场直接支持的任一时刻所能照应的独立图式的最大值，所以任何心理一功力的测验必须满足下面五个要求：

1.为了减少从知觉上激活图式的可能性，测验必须要对输入进行某种转换；

2.为了保证所有被试都能得到必要的执行图式和运算图式并且以最小量的心理一功力激活这些图式，测验必须就所要求的转换提供充分的先前训练；

3.测验必须利用那些要加以转换却已经充分学会了的单元，但这些单元不是测验的较多级的记忆组块的一部分，为了保证要求用一个并只用一个心理一能量单位来激活有关的形象图式，这样做是必须的；

4.测验决不容许用任何节省功力的策略或记忆术的策略来作出解答。因为这些策略就会使实验者所计算的单位数目大于被试实际激活的图式的数目；

5.必须使被试充满情感，以便将他的全部心理一功力运用到任务中去。

数字安置测验(Digit Placement Test)是满足上述要求的一

个实例(凯斯，1972年C)。在这个测验中，要求儿童执行的转换是把一个已知的数字(如4)放到一个数字序列的正确位置上，这个数字序列的各个数字的位置是固定的，并且从左到右的数值逐步增大(如，2、9、11)。实验之前，先要对儿童进行完成这一操作的先前训练，直到儿童能够迅速而成功地连续五次完成任务为止。在达到这个标准之后，让儿童坐在图3.1所示的装置前面。这个装置每次显示一个固定的数字序列，然后呈现需要安置的数字。为了保证被试熟悉测验使用的基本单位，我们把数字限制到20以下的范围内。另外，通过先前训练筛选出那些不能迅速而准确地标记这些数字的儿童。为了排除数字组成串的可能性，所选的数字序列要考虑能避免儿童常见的和容易认出的任何形式(如5、7、9)。为了防止儿童运用编组和背诵的记忆策略，当每个数字出现时，要求他们标出来，而且每当儿童一标出眼前这个数字时，就呈示一个新的数字。最后，为维持住开始时由仪器引起的强烈动机，对儿童的作业和心境(mood)给予适当的反馈。

当我们把这种任务给予6岁、8岁和10岁的儿童时，所得的平均数与根据这一理论所预测的数字是相符的。实际上，6岁组儿童能够正确安置的数字最大是3位，8岁组是4位，10岁组是5位，在这些临界点之后，每个年龄组分数的分布就与偶然预料的分数区分不出来了(凯斯，1972年C)。

数字安置测验并不是满足心理一功力测量基本要求的唯一任务。至今，已经设计出总共八个不同的测量，它们用了各种不同的格式和材料，并涉及到了听觉和视觉这两种感觉道。虽然没有一个测验能适合于所有的年龄组，但是这些测验都表现出接受它们的年龄组的假定常模〔伯蒂斯(Burtis)，1974；迪亚茨(Diaz)，1974；帕金森，1976；帕斯考尔一莱昂，1970；托塞恩特(Toussa-int)，1974〕。

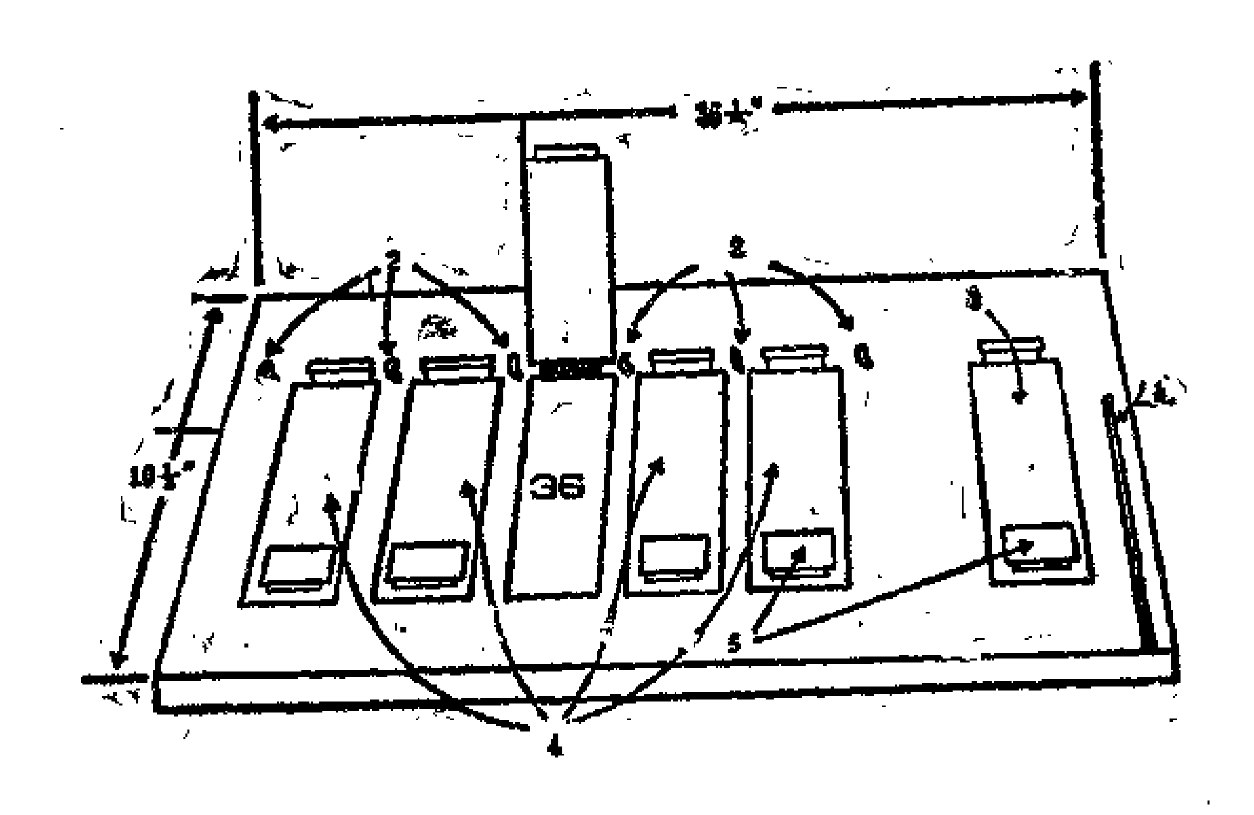


图3.1 测验用实验装置

1-记号；2-放东西的挂钩；3-隐藏最后一个数字的门；4-隐藏上升系列的门；5-门把手（根据Case的"新皮亚杰学派心理能力结构的证实"一文改编的，实验儿童心理学杂志' 1972，14，287—302）。

B.心理—功力与皮亚杰任务作业的关系

因为假定儿童的心理—功力是影响他建构和运用的执行策略复杂化的主要因素，由此推断，在心理—功力和皮亚杰任务作业之间必定有着重要的关系。已用几种不同的方法研究了这种关系的性质。

第一，心理—功力的测验与皮亚杰的运算，测验同时实施，两类测量的结果是相关的。当这样做的时候，无论在具体运算水平上还是在形式运算水平上显示的相关都是中等的（如0.4—0.6）〔凯斯，1977b；多尔，1976；德里博皮埃尔(DeRibeaupirre)，1975；

帕金森，1976；托塞尔特，1974〕。当通过回归分析排除了年龄影响时，那么相关系数就降低了(正如所预料的那样)，但是在统计上仍然是显著的〔凯斯，1977a，1977b；多尔(Dale)，1976〕。

第二，对皮亚杰任务作了修改，以至增加或减少这些任务的心理—需要，并把这些任务施于具有不同心理—功力的各组儿童。最近已由斯卡达马利亚报告了一个这类设计独特的研究(1977b)。她感兴趣的是儿童在既定的一组物体中摆出尽可能多的组合的能力。她向儿童呈示一套如图3.2所示的卡片，并要求儿童尽可能地组成以三张卡片为一组的许多个不同的集合，每一个集合总要包含各纵行(Column)中的一张卡片。大多数做对的儿童似乎运用了她比作里程表(Odometer)运转的策略，他们把每一纵行卡片看作象是里程表上的一系列数字。他们先从摆在右手边的卡片开始他们的循环路线依次一个一个地交换卡片，轮换完第一纵行上的全部卡片，如果第一纵行上的卡片全部轮换完，他们就向左移向下一纵行，交换这一纵行的第一张卡片；然后再次通过右手边纵行的轮换。按这种程序进行，就有可能用象里程表在其刻度盘上得出所有可能的数字组合完全相同的方式，得到所有纵行卡片的可能组合。在不同的纵行中，他们并不总是按同一方向进行的，即在某些纵行中，他们是从上到下进行的，而在另些纵行中则是由下向上进行的。但是，在任何一纵行中，他们总是有规则地从一端向另一端进行的。根据这种情况，斯卡达马利亚推断，在任何一列纵行中，儿童需要通过激活的一个独立图式来留意按什么方向进行，而对每一纵行的每张卡片就不需要激活一个独立的图式。据此，她得出结论说心理—需要应该等于卡片的纵行数目而与每一纵行的卡片数目无关。

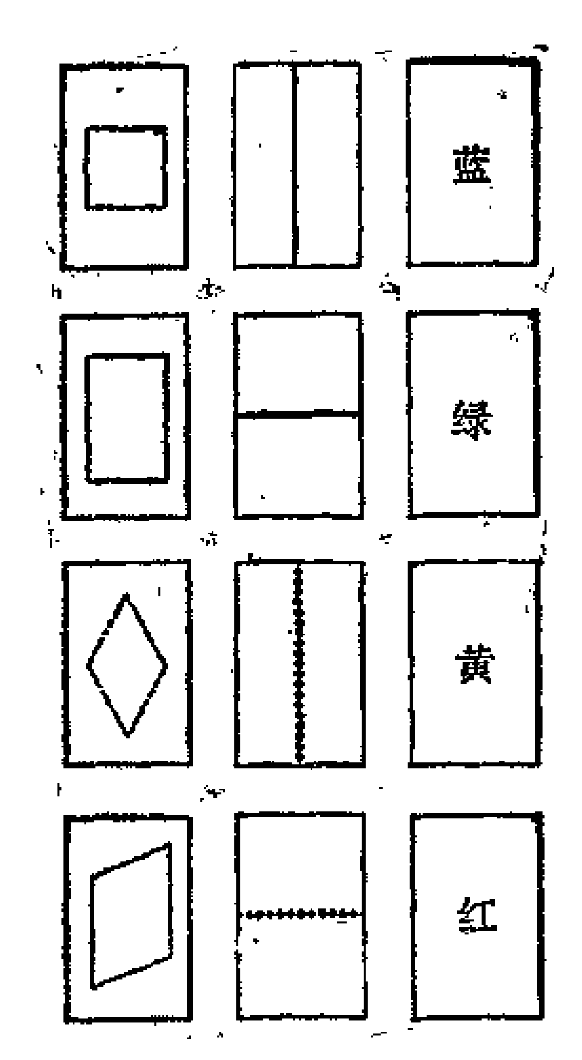


图3.2 在组合任务中呈现给儿童的一组卡片(卡片是用醋酸纤维素做成的,当它们互相叠成堆时形成了一种混合的刺激)

基于这种分析，斯卡达马利亚用纵行数目和每一纵行卡片目都不同的一组任务研究了三个不同心理—功力组的作业。对被试要加以选择以便使其心理—功力的值是4、5、7,在用纵行数目很少的题目进行练习之后，对每组被试都是有规则地增加纵行的数目,直到超过被试已测得的心理—功力为止。如表3.7所示，绝大多数被试都能容易地完成任务，直到纵行的数目超过他们现有的心理—功力为止。用修订过的“变量控制测验”(斯卡达马利亚1977c)、“水平测验”(帕斯考尔—莱昂,1969)、以及几种分类测验〔帕斯考尔—莱昂和史密斯(Smith) 1969〕都得出了类似的结果，

为了研究心理—功力与皮亚杰任务作业之间的假设关系，我们所使用的最后一个策略涉及到了教学运用。对几个不同年龄组的儿童，我们教给他们一个完成标准皮亚杰任务的策略，根据理解和运用这一策略所要求的心理—需要来预报掌握这一策略的最小年龄。最近一项教儿童学会控制变量的研究，对这种研究提供了一个很好的说明(凯斯，1974b)。控制变量的策略是一种比较简单的策略，至少在提前知道所感兴趣的变量并且这些变量的知觉特点比较明显的情境是如此。被试必须做的全部工作就是在所测的维度上找出一个具有正极值的物体(如一根长棍)，再找出一个是负极值的物体(如一根短棍)，然后进行检验，看看可以影响最终感兴趣的两个物体之间是否还有其他任何差别(如变弯曲)。在表3.8中对这一策略进行了分析，从中可以看出这个任务的心理—需要仅是3。

表3.7　在不同的复杂性水平上能运用组合策略的被试的百分数[a]

组的心理—功力	呈现的纵行数目						
	2	3	4	5	6	7	8
($N^4=15$)	100%[b]	100%	87%	15%	0%	c	c
($N^5=15$)	c	100%[b]	80%	73%	18%	0%	c
($N^7=3$)	c	100%[b]	100%	100%	100%	100%	0%

a.对斯卡达马利亚1977年所提数据的重新分析。

b.在最低的复杂性水平上，百分数界定为100。表示在这种水平上进行练习之后，没有策略的被试假定为不具有这种策略。

c.没有测验过。

表3.8　　变量控制的详细模型[a]

步骤或运算	具体的图式	符　号[b]
1．找出具有所测维度的正极值物体(如长)	(1) 相当于所测维度正极工作定义的运算图式(如对长度来说，如果物体伸出最多，那就称它为最长的)	ψ 测验维度(+)
	(2)c表示视野中物体序列的形象图式	φ 序列
2．找出具有所测维度其它极值的物体。	(1) 相当于所测维度其它极工作定义的运算图式(如，对长度来说，如果物体凹进去最多，那就称它为最短的)	ψ 测验维度(-)
	(2)c表示视野中物体序列的形象图式	φ 序列
3．检验看两个物体之间而不是所测验的维度之间是否有任何差别	(1) 表示所测维度的形象图式	φ 测验维度
	(2) 表示在两个物体之间来回扫描并分离出它们之间的任何明显差别的程序的运算图式	ψ 发现差别
	(3) 表示物体A的形象图式	φ 物体A
	(4)[c]表示物体B的形象图式。	φ 物体B
4．如果发现了差别就再回到第二步		
5．如没有发现差别，那就进行属性测验(即在感兴趣的属性，如变弯是否物体A>B)		

注：引自R.凯斯的(《结构和限制：在认知发展过程中某些机能的局限》，认知心理学，1974，6，544—573)

a．提问：已知一组具有属性X的多维物体(如变弯)，那么维度Y影响X的大小吗？

b．ψ=运算图式；φ=形象图式

c．由知觉场激活的图式，不必有心理—功力。

为了教给儿童这种策略，设计了一种教学序列，给儿童提供了前面所说的两种指导：(1)指导他们认识到他们当前使用的策略是不合适的；(2)指导他们如何建构一个新的策略。因为任务分析已经揭示了与这种指导相应的心理一需要也等于3(凯斯，1974b)。所以可以预言：7、8岁的儿童能够获得变量控制策略，并能把这种策略运用到皮亚杰所设计的那些真正的新问题中去。也可以预言：5、6岁的儿童是不可能做到这一点的。正如表3.9中所示，这两种预言得到了完全肯定的证实。后来在6岁组和7岁组的儿童中重现了这些结果，这些儿童被分成年龄差不多相等而心理一功力不等的两个组(凯斯，1977a)。用守恒经验(凯斯，1977b) 和分类测验(凯斯，1972a)，以及用严格说来不是皮亚杰的几个测验(凯斯，1974a，1975)，也都得到了类似的结果。

表3.9　完成两项要求变量控制的新任务的被试百分数

年　龄(岁)	教　过		未教过	
	任务1	任务2	任务1	任务2
7—8	80	80	20	40
5—6	0	0	0	0

C.场独立性与皮亚杰任务作业之间的关系

因为我们假设场依存性只是在含有场效应的内容领域里才影响被试的智力发展速度，这样我们就可以推断：在场独立性测验与含有场因素的皮亚杰任务之间，应当存在着一种重要的关系；而在场独立性测验与不含有场因素的皮亚杰任务之间存在着一种微弱的或不重要的关系。

正如威特金(Witkin)所界定的那样，场独立性的认知风格指的是被试克服他所面临的知觉场影响、或者把一个项目与它的背景分离出来的能力或倾向。评价被试场独立性程度的标准测量是棒框测验 (Kod and Frame Test)。做这个测验时，先把被试带进一个暗室，暗室里装有一根用发亮框架围着的发亮的棍，框架倾斜28°，再没有其他可供垂直判断的视觉暗示了，要求被试把棍真正摆成垂直。这类测验所显示出的个别差异是非常可靠的，并且长时间保持着明显的稳定性（威特金、古迪纳夫Goodenough和卡波karp，1967年)。

帕斯考尔—莱昂(1969年)在完成博士论文时，对一组9岁儿童和一组成人做了棒框测验，同时还做了另外几个场独立性测验和一套具体运算测量。他发现在场独立性测量与皮亚杰预先判断含有场效应的测量之间存在着显著的组内相关，而场独立性测验与皮亚杰判断不含有场效应的测量之间存在着较低的或不显著的相关。多尔最近也进行了一项有关的研究(1976)，他对一组10岁的儿童和一组13岁的儿童进行了一个场独立性的标准化团体测验、一组心理—功力测验以及几个皮亚杰的形式运算测验。他发现甚至在由心理—功力和年龄所引起的差异消除之后，场独立性的测量和皮亚杰的形式运算测量之间存在着显著的组内相关。尽管他们的设计不够全面，但是其他研究者也进行了相似的研究并得出了类似的结果〔凯斯，1973a，1974b，萨尼(Saarni)，1973〕。

D．经验与皮亚杰任务作业间的关系

正如前节所提到的，对经验与皮亚杰测验作业之间的关系已进行了大量面详尽的实验研究。大体上说，日内瓦的发展心理学家感兴趣的是自发发展的机制，并且所得资料表明经验只具有适当的作用(英海尔德等，1974年)，相比之下，北美的发展心理学

家感兴趣的是编序发展的机制，并且至少在最近的研究中，已经获得的资料表明经验具有巨大的作用(格尔曼，1969；利菲弗里和皮纳德，1972)。新皮亚杰学派的观点认为两类经验都能够产生真正的认知上的变化，并促成每种变化的基本机制是非常相似的。新理论的假设是：两类经验的差别只在于它们对被试的搜索指令系统和心理—功力的要求有所不同而已。

为验证这种假设早期所设计的实验研究主要想表明：编序发展的作用是很大的，而被试的心理—功力仍然限制着他所能建构的策略的最大复杂程度。这种研究的控制最完善的实例是前面引用的变量控制实验(凯斯1974b)。但是，还进行了许多其他的研究，都显示了很强的训练效果，并且这种效果的大小与被试的心理—功力之间表现出一种同样强烈的相互作用(凯斯1972a,1974a,1974b,1977a,1977b)。

最近，进行了一种尝试，表明这种理论不仅能预测经验在高度编序学习情境中的作用，而且还能预测经验在比较自发性的学习情境中的作用。虽然这方面的许多研究正在进展中，但是包括守恒范例在内的三项最新研究已经完成。在第一项研究中，使被试面临由利菲弗里和皮纳德所使用的那种高度编序经验，结果表明通过这种训练，被试完成任务的心理—需要被降到2，在心理—功力是2或稍高一点的被试中，有50%的人表现出有所提高，而在心理—功力是1的被试中没有一人表现出有任何提高(凯斯，1977b)。在第二项研究中，向儿童提供的经验类似于所使用的一般类型的材料但有一个基本方向不同。尽管为儿童提供了进行数量比较的机会并使儿童能接受到关于他们判断正确性与否的反馈，但是从不引导他们通过表3.4和3.5中所述的一系列心理步子，结果形成守恒策略的心理—需要仍然是3。心理—功力是3的被试，他们能够独立通过这个心理步子序列并能在没有实验者任何帮助

的情况下去建构守恒原则。相比之下，心理—功力是2的被试则什么守恒原则也没掌握。无论是在第一项还是在第二项研究中，心理—功力的作用和年龄的作用被混淆了，但是，正如表3.10所示，在第三项研究中获得了相同基本类型的结果，在第三项研究中对年龄因素进行了实验控制。

表3.10　通过两项守恒任务被试的百分数教学的类型

心理—功力	直接的	间接的
3	100	50
2	57	13
1	0	0

注：引自R.Case的〔"在认知发展中阶段过渡的过程"(最后一篇报告，R01HD09148-01、MMHCD，1977)Berkeley、Calif·加利福尼亚大学，1977〕

E、场独立性和心理—功力之间的关系

在新的理论体系中，可以用完全象分析皮亚杰任务作业一样的形式来分析场独立性的测验作业，找出形成正确答案所必需的一系列心理步子。根据帕斯考尔—莱昂的观点，这种分析揭示出正确解答，不仅要求被试运用一个由教学而不是由知觉场所指点的执行图式，而且如果他们所用的一个执行图式是直接由知觉场激活的话，那么，就要求他们去协调比他们所需要的图式多得多。(帕斯考尔—莱昂，1969)。情况既然如此，帕斯考尔—莱昂就把威特金和他的同事在场独立性测验中发现的发展上的增长，归因于众所周知的心理—功力发展上的增长。与威特金不同，他把场独立性的年龄内的差异归因于场敏感性本身的差异，还归因于要尽量把心理—功力动员起来的这种意向上的差异。

在最近的两项研究探讨了场独立性和心理—功力之间的关

系。在第一项研究中，对一组二年级的学生实施了一套心理一功力测验，和一套场独立性的测验(凯斯和格洛 伯 森，1974)。根据这种新理论，这两套测验负载着不同的但又适当的因素。在第二项研究中，试图确定这两种因素之间的关系是归因于心理一功力的动员意向上的差别，还是归因于心理一功力本身的差别。(格洛伯森，1976)。对同一所学校同一个年级的一组儿童进行了一套类似的测验。但是，这次把测验修改得可以同时估计儿童的瞳孔扩张。而已经有实验证明可以由瞳孔扩张的增加来可靠地预测心理努力的增加(赫斯，1972；卡尼曼1973)。通过协方差分析消除瞳孔扩张的作用之后，场独立性和心理一功力之间的关系确实消失了。但是，也必须指出：(1)场独立性和心理一功力之间的关系是曲线的，而不是直线的；(2)测验分数和瞳孔扩张的协方差几乎全部表现在心理一功力测量方面。就场依赖性测量本身来说，如果具有场依存性的被试确实因心理一功力动员意向低而获得的分数也低，那么场依存性的作用就应该是最大的；完成测验不佳而瞳孔扩张又比较小的被试，就没有任何要动员心理一功力的意向。目前，所能得出的最保守的结论是，心理一功力和场独立性之间的年龄内关系的性质还需要进一步加以阐明。

总之，到目前为止虽然所进行的研究数量不多，但是基本理论假设已经得到很多实验的支持。实验已表明心理一功力是按照最初假设的尺度随着年龄而线性增长。另外，它们还表明，心理一功力、场独立性和经验这三种因素都对儿童智力发展速度独立地起作用。目前还确定不下来的是：场独立性和心理一功力之间这种基本关系的实质。

Ⅵ. 关于新皮亚杰学派的教学理论

本文的第三节曾指出，如果皮亚杰的理论是正确的，那么任何基于发展的教学理论必须解决两个问题：第一个问题是如何使儿童运算结构的发展达到最佳化；第二个问题是如何使教学内容适合于儿童已经达到的结构。从一般的分析水平上看，对上述两个问题的回答是清楚的。第一步是分析作为教学内容的某个学科的基本结构；第二步是评价儿童在那一学科中的现有实际功能水平；第三步是设计适于儿童现有功能水平的教学活动，或者是为了促进这种功能水平向更高一级的水平过渡(如果那是目标的话)或者是为了选定符合现有功能水平的教材(如果那也是目标的话)。

因为刚刚阐述过新皮亚杰理论它仍保持着皮亚杰理论的基本概念框架，所以新皮亚杰理论所提示的一般教学步骤同皮亚杰理论所提出的一般教学步骤是一致的。另一方面，由于新理论是一种功能性的理论，而不是一种结构性的理论，因此，当试图把这种一般步骤变成实践时，就要排除所遇到的许多困难。在这部分中，将重新考虑上述三步中的每一步，目的在于说明当用新皮亚杰理论观点来探讨这些任务时，如何克服和减少所遇到的实际困难。

A.促进运算结构的发展

1.第一步：结构分析。在传统的皮亚杰学派理论中，由于两个原因，对常规学习任务难以进行结构分析。第一个原因是所要分析的结构类型是逻辑结构，而且在大多数常规学习领域中，逻辑结构的作用和关联并不明显；第二个原因是所要分析的类型不

是作业分析而是能力分析，因为学生的作业是学生能力的一种不确定的指标，还因为皮亚杰没有提出从一种分析类型到另一种分析类型的一般方法，因此，很难知道分析从何处着手或者如何进行。

在新皮亚杰学派的理论中，上述两方面的问题都不存在。第一，所要分析的结构类型是一种执行结构，并且在大多数学科领域中这些结构的重要性是明显的；第二，因为，一种执行结构的功能精确地控制着被试在既定任务方面的作业，所以执行某一任务所需结构的特点可直接从熟练的被试所完成这一任务的作业中推断出来。虽然在对一种任务进行执行结构分析时，并没有什么必须遵循又必不可少的步骤顺序，但是下而的一系列步骤是极为有用的，并且在一般情况也是足够的了。

1.1 第一步是确定所要完成的任务的目标。这一步只要通过考查向儿童提出的问题就可以实现了。例如，在守恒任务中，问儿童："哪个多？"于是，基本执行图式所定向的目标是确定相对数量。

1.2 第二步是描绘出一个成功的被试达到这个目标所经历的一系列步骤，进行这一步骤的一种很有用的方法就是一个人亲自去执行标准任务，并列出达到目标所要通过的一系列操作。当用这种内省法时，首先有助于描绘出这个人步骤的一般序列，然后根据后来的尝试列出这一般步骤的每一步里他所执行的子操作顺序。例如，如果被试企图描述他的双维定量的策略，那么他要注意的第一件事是他看一下高度和宽度。然后执行这个任务并监控对每个维度进行实际比较的子心理步骤。

在某些任务中，运用内省方法是困难的，因为对所提问题的回答似乎没有经过任何心理步骤就很快想出来了。例如，守恒任务本身就是这种情况。被试只"知道"这两个量必须仍然相等，而

不必了解是如何知道的。在这些情况下，一种很有用的方法就是从纯理性的观点去探讨问题从而设计出一系列在逻辑上是必不可少的步骤去确证如此明显的答案。帕斯考尔—莱昂正是使用这种方法进行了表3.5所描述的守恒执行分析。

1.3 描绘出找到正确答案的一系列假设步骤之后，下一步就是要把这一系列假设步骤与被试的实际作业进行比较。进行这一步的一种有效方法就是要注意有经验的被试在执行任务时所表现的动作和眼动的顺序。正如已经描述的那样，在被试正在完成任务时，真实地拍下他们的眼动照片，这种方法已极准确地用于守恒任务中(奥布赖恩和博斯马，1971)。但是，对大多数与教学有关的问题来说，进行一个非正规的分析就足够了。

另一种有效的方法是与熟练完成任务者交谈，问他是如何完成这项任务的，或者要他描述自己在实际完成任务时的思考内容(纽厄尔和塞蒙，1972)。这种谈话记录的分析，对那些年龄已到能充分意识到自己思维过程的被试和对那些实实在在要花相应时间的任务来说，是特别有用的。

1.4 根据被试的实际作业确认出一系列假设步骤之后，就需要修改这些假设的步骤，并且再次重复1.1到1.3的步骤。如果还存在任何双关意义时就需要编一个新问题，在这个新问题中参数改变了，以期望被试在运用第一种可能的策略时给予一种回答，而在运用第二种可能的策略时给予另一种不同的回答。例如，如果问题是正在执行守恒任务的儿童是真正注意液体转换的性质，还是只根据两种液体的最初状态作出回答，那么就可改变一下所执行的转换类型并看到他们的回答是否相应地改变了〔利菲弗里和皮纳德，1972；沃森(Watson)，1968〕。

1.5 最后，对一系列假设的心理步骤确实与熟练操作者(或者至少在某些子步骤上是熟练的操作者)所用的心理步骤相一致而

感到满意之后，偶尔也想检验一下关于完满性的描述。这样做的一种有效方法类似于计算机模拟所用的方法：辨认出描述初学被试的步子顺序，并让他只按教他的步子去做。如果碰到必须对所教步子加以补充或修改才能保证被试的行为符合预期的要求的情况时，那么对原来的策略描述就必须做相应的修改或补充。

2.第二步：学生现有功能水平的评价。在传统的皮亚杰理论中，由于存在着不受控制的作业因素而给儿童现有功能水平的评价带来困难。这些非控制因素的存在显然是由于下面两种情况造成的：一种情况是，在假定能测出有关的基本结构的任务与任务之间存在着低度的交替相关；另一种情况是，在假定测出相同基本结构的任务与任务之间存在着一些不可解释的发展阶段。如果存在着上而这两种现象，那么就不可能确定所得的测验结果，究竟是由于一种基本逻辑结构的存在或不存在所造成的，还是由于个体能够或不能够对付那种尚未查明的、又为任务所固有的作业因素中的一个所造成的。

如前所述，概括起来说，新皮亚杰学派理论是够准确地阐明大多数皮亚杰作业所固有的作业因素，以至任务间高度交替相关是否存在、儿童发展阶段是否存在，这都能预测到。(帕斯考尔一莱昂1969，斯卡达马利亚1977b；托塞恩特，1974)。如果确实如此，那么，就有可能把某一被试的失败由于下述任一因素利用某一执行结构而归咎于下面的任一因素：(1)缺乏适当的心理一空间(通过测量这个建构一个异常低的分数来表示的)；(2)对有关的知觉暗示的异常敏感(通过场依存性测验一个异常高的分数来表示的)；(3)缺乏所测查领域的具体经验(通过以上两种测验的平均分数而又不能完成任务来表示的)。

从特殊教育的观点来看，确定造成特殊儿童困难的原因所在，是大有希望的。〔巴乔(Bachor)，1976；伯阿维拉(BeAvila)和哈瓦

赛(Havassy),1974)。当一个特殊儿童的问题是严重的并表现在多种多样的内容领域里时,那么,了解其病因是很有用的。这有助于确定现实的目标和制定有效的矫正教育方案。然而,从最正规教育的观点来看,新理论的威力并不在于它能诊断出特殊儿童不能运用特定执行结构的原因,而首先在于它重新界定了有关结构的性质。正如后面各节将要说明的那样,儿童在一个特殊的背景下运用一个不适当的执行结构的原因常常是无关紧要的;而重要的事情是:儿童实际运用的是什么样的执行结构。为了设计有效的教学,这一点必须要评估到,也必须要了解到。

因为评估的最重要的内容是在所测查的领域或任务中儿童现有的执行策略,所以评估的方法与前节提到的方法是相同的。所不同的只是人们必须要去分析的是导致错误答案的策略,而不是导致正确答案的策略。

实际上,评估不能完成任务的学生的策略比评估能够完成任务的学生的策略要困难得多,这可能由许多原因造成的:这是因为人们自己使用着正确策略,很难设身处地站在儿童的立场上去设想出一个导致儿童答案的策略。因为在不能解答问题的儿童的表达能力常常要比能解答问题的儿童差得多,所以要获得关于他们的内省或论证的资料也会是更加困难的。最后,因为人们的目标总是教正确的东西,所以难以离开这个中心,难于把儿童的答案看成决非"错误"的东西(或者,充其量某一特定成分是有缺陷的)。因为分析一个错误策略比分析一个正确策略更加困难,所以能把分析错误策略的问题转变成分析正确策略的问题常常是有益的。至少有两种方法可以达到这种转变。

2.1 一种有效的方法是确定一个能让儿童作出正确回答的问题,然后循环通过从1.2岁到1.5岁,好象这真的是这个任务的目标。例如,在守恒实验的例子中,如果问题是:"两个量杯看起来

哪个好像装得更多一些?"那么儿童的回答将是正确的。如果向成人提出这个问题，那么他们采取的策略与儿童使用的策略非常相似，并且可以用同一种方式去分析这种策略。

2.2 与第一种方法同样有效的另一种方法是假设被试没有机会得到他好象忽略了的某一极其重要的信息。通过向成熟的但又没有这一信息的学习者提出任务，人们可以再一次假设达到这个目标的程序，并把这种假设的程序与实际使用的程序相比较。例如，在守恒实验的例子中，儿童的动作明显地表现出似乎对最初状态或转换都不知道。如果人们仅向成人呈观两个直径不相等的量杯并且问他们哪一个装得多一些(或者甚至是人们想象一下自己将如何去解决这个问题),那么，要了解儿童策略的基本原理和详细说明策略所必需的心理步子，就更容易了。

3.**第三步：教学设计**。在传统的皮亚杰学派理论体系中，把阶段过渡描述成为一个具有五个时期的过程。在第一个时期中，儿童完全依赖一些实用而孤立的图式；在第二个时期中，儿童获得了一种新图式，这个新图式就成为一个冲突的潜在根源；在第三个时期中，儿童体验到两种图式之间的冲突；在第四个时期中，儿童根据冲突的解决建构一种更为高级的结构；最后，在第五个时期中，儿童巩固和扩大他们的新结构。如果这种一般描述使人非相信不可，那么，这种描述的困难就在于：给那些有志于把运算结构的获得过程最佳化的人们所提供的指导是不充分的，因为这种一般描述并没有阐明影响阶段过渡过程的任何成熟变量和经验变量。

在新皮亚杰学派的理论体系中，上述这种困难大大减少了。详细阐明了一种成熟变量，也就是心理一功力。另外，还详细阐明了三类经验变量，其中的每一类变量都是可能控制的。这些变量是：(1)儿童所处情境的熟悉性；(2)儿童必须照应的线索的鲜

明性；(3)儿童必须协调的信息项目的个数。然而，从新皮亚杰学派的观点来看，能使儿童获得一个特殊执行策略的最有效方法是引导儿童通过由皮亚杰划分的这五个时期，同时一定要最大限度地提高情境熟悉性，提供尽可能鲜明的有关线索，并且把进行有关图式协调所要求的心理一功力减少到最低限度。虽然，儿童心理一功力的大小限制了他所获得的策略的复杂性，但是，就使用这种一般方法会保证任何既定的心理一功力水平来说，所能获得的策略的复杂性要比在正常情况下大得多。简言之，它将保证一个低的心理一功力将尽可能少地干扰学习。

为实现上述的目的，下面的一系列程序是有效的：

3.1 第一步是要建立一个被试可以从中对自己目前使用的策略的有效性进行评估的范式。在这一步中，必须十分注意把情境的熟悉性和线索的鲜明性增强到最大限度，而把认知的复杂性缩减到最小限度。例如，假定人们的目的是教儿童一种控制多变量的复杂策略。在这种一般用于评估儿童这一领域的技能的范式中，与那种只操作一个变量而让其它变量保持不变的基本要求不同，至少有三方面的困难(斯卡达马利亚，1977C)。第一个困难是因变量和自变量常常是被试不熟悉的；第二个困难是自变量彼此之间没有明显的差别，而且任何一个变量的变异常常是难以觉察的；第三个困难是须操作的变量总数是相当大的。如果人们的目的是教学而不是评估，那么第一步是要除掉任务中的上述这些困难，并且编制出某种供被试测试自己所用策略的有效性的程序。图33.展示了一个符合这些要求的范式〔凯齐斯(Kenzie)，1970〕。任务的指导语如下："这是一台天平，你知道它是怎样工作的吗？好，你能用它来告诉我是哪种棒重一些吗？是灰色棒重一些呢，还是银色棒重一些呢？"

要注意这个范式是如何消除了上面提到的那些困难的。因变

量和自变量都是为儿童所熟悉的变量，唯一新奇之处是这些变量所取的特定值(即所使用的特殊类型的材料和它们的具体重量)。两个自变量，棒和块料是有很大区别的，每一个变量的变异是容易觉察的(是用不同的颜色标记的)。最后，自变量的数目被缩减到最小限度，即只有两个自变量。

还要注意，为了给认知冲突安排一个阶段，这个范式引进了另外一个特点。被试自发使用的策略的有效性，可以用一个简单而直接的程序来加以评估。一旦儿童做完一次测试并得出结论，他就可以把棒从块料中抽出来并直接称其棒的重量。

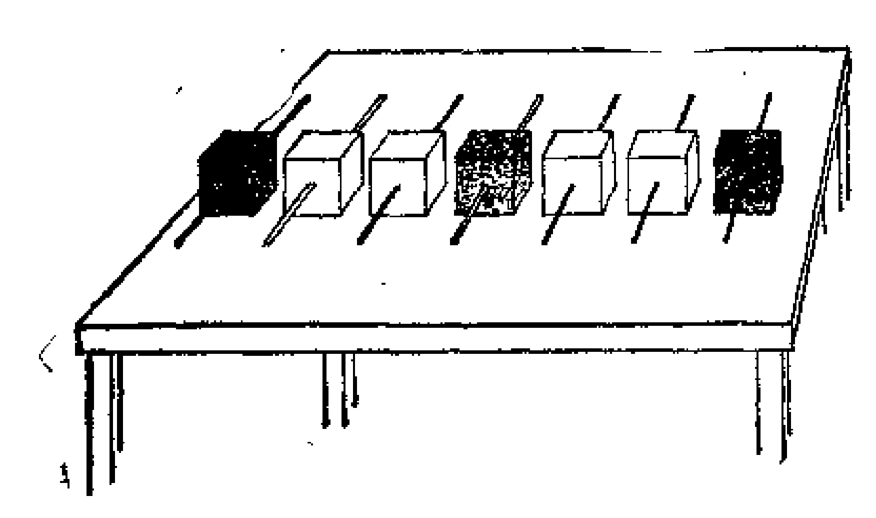


图3.3　训练儿童控制多变量所用的棒和块料阵列。

3.2 在设计出儿童能够评估他现有策略有效性的范式之后，下一步就是要去证明这个策略的不足。如果儿童的最初策略已经在评估阶段正确地诊断过了，那么只要选出一个合适的问题一般就能完成这一步。继续谈这个控制变量的例子：年幼儿童的自发性策略一般是挑选出他们看到的前两根不同颜色的棒，并且不厌其烦地称这两根棒的重量来检验分别嵌放这两根棒的块料的类型。只要按照下面的方式安排块料：离被试最近的两个块料是用不同的材料制成的，并且较重块料嵌放着较轻的棒，把效应弄混，这样就可以显示出这个策略已无效。当做完这一步以后，儿童都

一律得出结论说铝比黄铜重，而当他们移动块料时却发现并非如此，他们大吃一惊。

3.3 在向被试证明他的最初策略无效之后，下一步是帮助他发现为什么会是这样的理由。① 必须再一次考虑图式的熟悉性、线索的鲜明性和认知复杂性所产生的约束力。如果儿童自己没有看出他失败的原因，那就要用手势或言语指出关键的特征，然后再用简单而具体的句子和生动的言语反复说明这关键特征所起的作用。对块料问题可以作如下的说明："你怎么认为你受骗了呢？你不知道吗？摸一摸这些块料，你受骗是因为这个块料是这样的重，以至它使天平下倾，这就使得银色棒看起来比较重，虽然银色棒实际上并不是那样重。

3.4 在对学生指出他现有的策略为什么无效之后，下一步是要促进他建构更恰当的策略。一旦儿童理解了他最初失败的原因，他就会自然地建构他的策略。然而，对儿童来说这一步可能是太难了，以至他独自完成不了，因为这种策略的重建活动可能是一种复杂活动。

3.5 在教会儿童一种新策略去克服他原有策略中固有的困难之后，最后一步是巩固和扩展这一策略。完成这一步的基本方法是进行练习，通过定期改变任务性质得到加强。当儿童运用新策略变得自动化了的时候，执行这一策略所需要的注意(心理一功力)就愈来愈少。所以儿意愈来愈多地注意其它的事物。而那些在第一步（1）从任务中删掉的复杂情况，现在又重新被逐步引入了。每引进一个新成分，再"退回"到最简单的问题形式上来可能是有帮助的。但是，如果所增加的成分分别用a、b、c、d来表示，那

① 这一步似乎并不是皮亚杰的五个过渡时期中的一个时期，但是它完全符合于在第四节中所描述的阶段过渡的展开模型，从教学的观点来看，这一步无论如何是非常有意义的。

么按下列顺序排列转移活动可能是有用的：a,a+b，a+c，a+b+c,a+d,a+b+d,a+b+c+d。当按这样的顺序去完成变量控制的任务，变量控制所得到的结果是非常令人兴奋的(凯斯，1974b)。虽然没有一个被试在开始时表现出具有复杂的变量控制策略，而在程序结束时，绝大多数8岁的儿童完成任务的情况比未经指导的15岁儿童的平均水平要好一些。

B.教学内容要适合学习者的运算水平。

在传统的皮亚杰学派理论体系中，"教学内容要适合学习者的运算水平"这个用语有两种意思：一种意思是修改给儿童所提的任务，使儿童没有必要用高于他们现有的运算功能水平；另一种意思是，向儿童提出的任务一般要求更高的运算功能水平，同时教儿童某种完成任务的具体办法和决窍，以便排除这方面的困难。例如，设计出某种简化的办法并且把它教给儿童，那么，某些要求用形式运算来完成的任务，实际上就可以用具体运算来解决(凯斯，1974b)。

在新皮亚杰学派的理论体系中，情况就更为复杂一些。因为儿童的运算水平既可以从他当前使用的策略去界定，也可以从他已经达到的心理—功力水平去界定①。"教学内容要适合学习者的运算水平"这同样的一句话，实际上有三种意思：第一种意思是，使任务符合学习者现有的策略(上述第一种意思)；第二种意思是，教解决复杂任务的简单策略(上述第二种意思)；第三种意思是，教会儿童用复杂策略解决复杂任务，但教的方式是，决不让学习顺序的主观复杂性超过被试现有的心理—功力。就后一种意思来

① 利菲弗里(1976)和帕斯考尔—莱昂(1976a)之间的不同意见，也就是说关于利菲弗里的学习程序是否能导致"真正的"运算发展，可以归结为在这两种界定中选取哪一种界定之间所表现出来的差别。

说，前节所述的那套程序，已经构成了一门使教学内容适合学习者的运算水平的工艺学。然而，即使按前两种意思中的任何一种来用“教学内容要适合学习者的运算水平”这个词语，前节所描述的那套程序，仍然构成合乎要求的基本工艺学的核心。

要了解为什么会如此，首先考虑修改教学任务这个目标，教学任务才不会要求运用比学习者现有水平更为复杂的策略。为了做到这一点，第一步是重新设计这个任务，使它的形式最简单、最为学习者所熟悉(参见3.1)；第二步是把这个任务交给几个不同年龄水平的儿童去做，以确定第一次使用正确策略的年龄(参见1.1—1.5)；第三步是设计一系列练习，以便把这一年龄水平的儿童从完成经过简单化、熟悉化了的任务引导到完成原来要解决的任务(参见3.5)。从上面所说的可以看出：这里所必须使用的这套程序，同为了诱导运算发展所必须使用的那套程序是相同的，所不同的只在于：重新安排了程序的应用顺序，删去了分析错误策略的程序(参见2.1)和修改策略的程序(参见3.2—3.4)。

下一个考虑的目标是发现和教会一种比解决复杂任务通常运用的策略更为简单的策略。在此情况下，第一步是分析熟练完成任务者所用的策略(参见1.1—1.5)；第二步是分析最初有关的年龄组所用的策略(参见2.1)；第三步是用熟练完成任务者所常用的，并经过适当修改的策略，或它的替换策略来进行实验。目的是要了解是否能发现一种策略，它既和一些熟练完成任务者所用的策略一样的合适，又不比一些完成任务的生手用的策略更为复杂。在发现或者构成这种策略之后，最后一步就是教这种策略(参见3.1—3.5)。再一次可以看到：所必须使用的程序与前节所描述的那些程序是十分相似的。唯一的差别是：必须额外增加一个步骤，在下节中将介绍一个例子说明这个额外的步骤是如何加进去的。

总之，由新皮亚杰学派理论引出的教学措施与由传统的皮亚杰理论所引出的教学措施在一般形式上是相同的。但是，因为新理论提出了结构获得和结构应用过程的详细模型，所以在实际试用这种一般程序时所遇到的困难就大大减少了。当目的是促进运算发展的时候，鉴于所分析的结构是可以从被试的作业中直接推出来的，所以，结构分析就变得比较容易了；鉴于所要评估的实体只是某一被试在最初有关的任务方面所使用的策略，因此个别评估就变得比较容易了；最后，鉴于明确阐述了把儿童从一种策略引向下一种策略所要控制的作业因素，因此，教学设计就变得比较容易。当目的是使教学内容适合学习者的发展水平时，在每一阶段上所用的程序和所得到的效益也是很相似的。所不同的只是人们希望删去分析和修改错误策略的步骤，或者增加发明新颖而简化的策略步骤。

Ⅶ．从理论到实践：新皮亚杰学派对课堂任务的分析

正如前节所述的，在最近关于发展方面的文献已经报告过，大多数促进复杂执行图式获得的程序已经或明或暗地应用于成功的训练研究上。但是，到目前为止，还未显示出这些程序可用于比较一般的训练情境中。我们已经知道在心理学实验室与课堂之间有很大的脱节，这似乎是要完成的一项重要任务。

我们举一个所谓的缺少加数的问题来说明课堂任务。这项任务要求儿童看等式4＋□＝7，并填上所缺的数。尽管这项任务在表面上看起来很简单，尽管它是大多数正规的小学一年级课程的一部分内容，但是事实上它是相当难的。大多数一年级的教师或是承认他们在这一任务的教学上是不成功的，或是说他们不再去设法教这项任务了〔奥哈拉(O'Hara)，1975年〕。另外，尽管儿童对这

个任务的回答是错误的，但却是高度一致和系统的。那么，作为对新皮亚杰学派方法效力的检验，这项任务似乎是一项好的任务。下面将论述如何把前一部分所述的原则和方法应用到设计一种改进教学的任务中去。

A.第一步：结构分析

要确定缺少加数问题的难度，第一步是对熟练成人在解决这项任务时所用的策略进行分析。假设在等式中所用的数字相当大，那么成人似乎是依次通过了以下的几步：

1.从左向右辨认符号；

2.注意到所要求的数是两个加数中的一个；

3.决定必须从和中(也是已知的)减去已知的加数；

4.注意已知的加数值，贮存；

5.注意和的数值，贮存；

6.从和中减去已知的加数。

在传统的皮亚杰理论体系中，上述一系列步骤中最难的步骤大概是第3步，因为为了完成这一步，被试必须使用一种在心理上与另一种运算(加法)作用相反的运算(减法)。正如我们所记得的那样，执行这种可逆性运算的能力是具体运算阶段的标志。

在新皮亚杰学派的理论体系中，至少在学习时期，这一步大概同样是最困难的。为了进行这一步，假定被试必须事先对加法和减法之间的可逆性关系有所了解，还假定被试需要的心理—功力是3—5，这要依所提供的证明的简单程度而定。

最后，无论在哪一种理论中，都可以料想到，如果人们教儿童一种不需要求助于可逆性的策略，那么较小年龄的儿童就应能解决上述任务。满足这种要求的一种策略与店员兑换零钱时所用的"点数"程序相类似：

1.从左到右读题目，寻找运算符号(+)；

2.查看加号两边的符号；

3.注意到所要相加的一个数是已知的，而另一个数是未知的(用方格表示出来)；

4.找出等号并看等号另一边的数，贮存这个数(7)；

5.从已知加数之后的那个数开始计数，每次递增一个数(5、6等等)。每点出一个数，就作一个记号(如一个手指)；

6.当贮存总数达到(7)时，就停止；

7.计算出所做记号的个数并把它填入方格中(3)；

8.(可供选用的步骤)从左到右重读符号，根据人们对数的知识的了解核对语言表达，确实证明答案是对的(如：4+3=7，对吗?)。

9.如果这个表述正确，就停止。如果不正确，再回到第四步。

尽管上述策略是由纯粹的理性分析得出来的(参见1、2)，但是，也有实验证明：与成人相比，学会了解这一问题的儿童倾向于运用这种策略〔格罗恩(Groen)和波尔(Poll)，1973年〕。

B.第二步：现有功能水平的评估

当儿童不能正确解决上述问题时，他们中的大多数人一般容易犯下述两种错误中的一种错误。那些没有做过加法训练的儿童或者学习迟钝的儿童，他们都倾向于把4或7作为答案。那些做过相当多的加法训练的儿童和学习正常的儿童，都倾向于把11作为答案。只要知道了他们犯了哪一种错误，就很容易诊断出所用的策略了(参见2.5)。如果教学只是在所给的位置上写出一个数字，那么，上述第一种答案将是对的，如果问题4+7=—，那么上述第二种答案就是正确的。据此，第一种策略可以描述如下：

1.看符号，辨认出其中的一个；

2.在方格中写出这个符号。

第二种策略可以作如下描述：

1.看第一个符号，贮存；

2.数出那么多个数的东西(如手指)；

3.看第二个符号，贮存；

4.数出那么多个数的东西；

5.数出前面所数出那么多个数的东西的总数，贮存；

6.在方格中写出这个数字；

7.(可供选用的步骤)进行核对，看看两个最小的数相加是否能得出那个最大的数(7+4=11,对吗?)；

8.如果对，就停止。如果不对，再回到第一步。

即使随便检查，也可以看出错误策略的复杂性与第二种正确策略的复杂性是很相似的。

C.第三步：教学设计

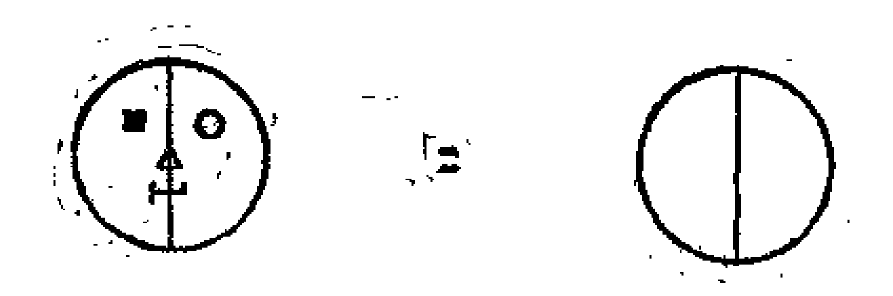
人们根据一个目标可以设计出一个教学序列把儿童从任一不正确策略引导到任一正确策略。为了作一说明，我只考虑已诊断出当前策略为第二种策略(即加法)的那些儿童的例子，对他们来说，其目标是教两种正确策略中较简单的一种，即在他当前所达到的策略发展水平上进行教学。读者如对改进功能发展水平的序列(即教更复杂的"可逆策略")感兴趣，可参阅拉姆的文章(1977年)。

3.1 教学设计过程的第一步是创设一个范式，其中儿童有某种评估他当前策略的适合性的简单方法。对这一步来说，许多个范式都是同样有效的。但是我在实际实验时所使用的是下面的一种范式：

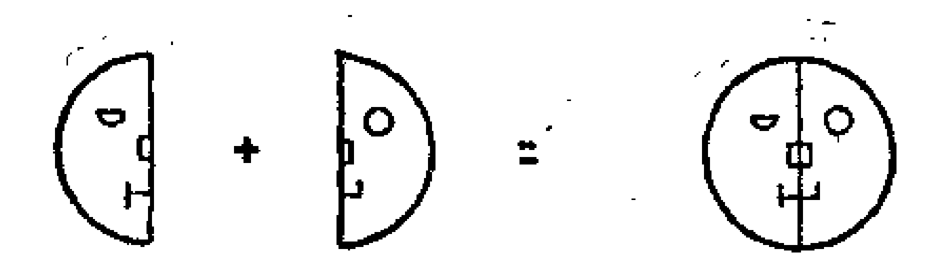
这些脸看起来相同吗?这个符号(=)表示它们是相同的。



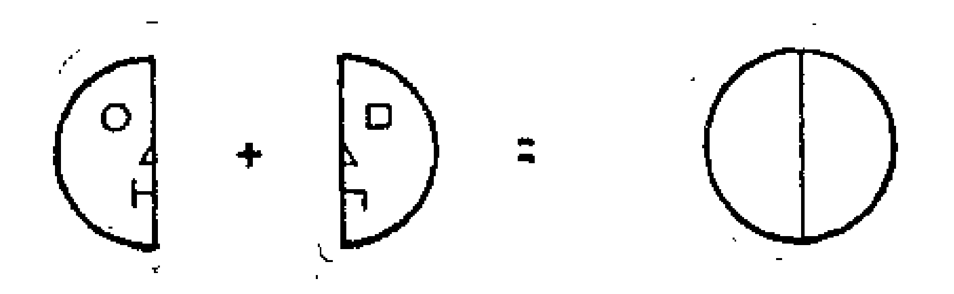
这里还有两个脸，你能使这个脸和那个脸相同吗？画上其中的一些图形，使这边的脸和那边的脸恰好相同。



这个符号（＋）是表示把这两个半边脸合在一起。你能看出当我们把这两个半边脸合在一起时所得到的一个脸和这边的一个脸相同吗？



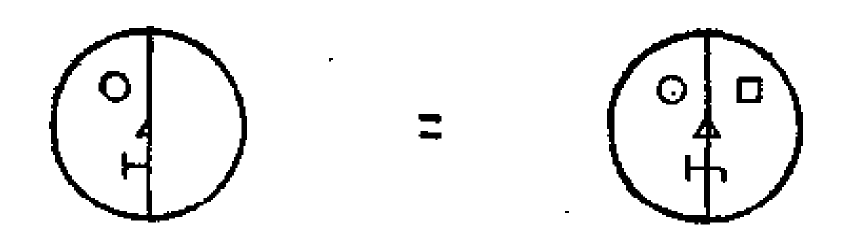
记住，这（＋）是说把这两个半边脸合在一起。现在你能使这一边脸和那一边脸恰好相同吗？



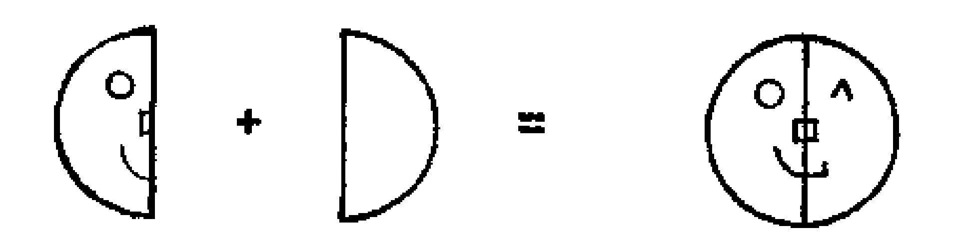
注意，这个范式符合前面部分阐明的全部标准，它是儿童所熟悉的。有关的线索是明显的，所有与使等号两边相匹配这一目标无关的要求，如计算和读数字的要求，都被删除了。还要注意，反馈的形式是这样的，使得儿童亲自看一眼就能说出他是否已达到了目标。

3.2　第二步是向儿童提供能表明他所选定的策略是不合适的那些问题。可以体现这种策略的各种任务如下：

记住，这个符号(=)表示使两边相同。画上其中的一些图形，使这边和那边恰好相同。



记住，这个符号(＋)表示把这两个合在一起，现在你能使这一边和那一边恰好相同吗？



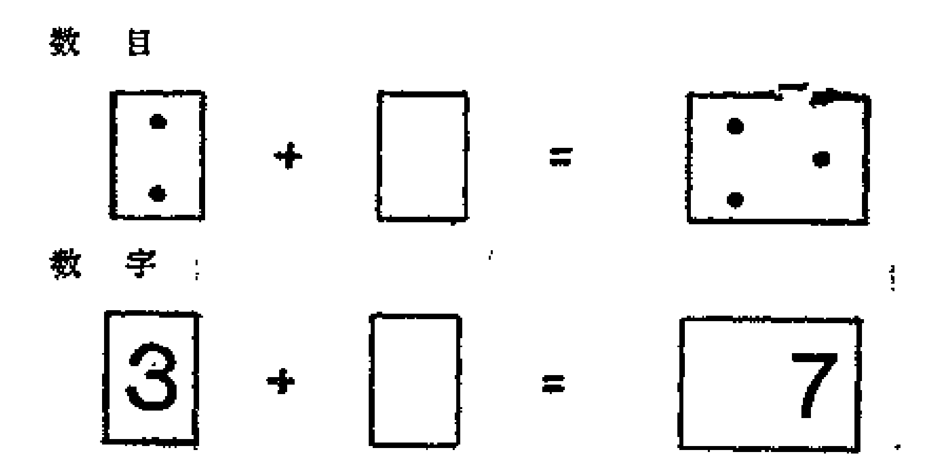
3.3　第三步是向儿童说明为什么他所选定的策略是不正确的。例如，当一个儿童使用了不正确策略时，就对他作如下的说明："记住，这个符号(＋)是说把这两个合在一起，以至使这两个半边脸合起来与那一个完整的脸相同。当你把这两个合在一起时，你能得到和那个一样的脸吗？"

注意这是如何满足了前面部分中所述的那些标准的。它吸引儿童去注意他们忽视了的有关线索(＋)，并用简单、简短和具体

这些词来解释他们的回答为什么是不正确的。

3.4 第四步是提供一个正确策略的模型。由于这种脸的范式涉及到很具体的材料，因此，在这个阶段上就不会是很需要的了。但是，大体上人们能把儿童的注意吸引到将完整的脸型分解为几部分的方法上，并核对每个部分以确保等号的两边相同。

3.5 最后一步是巩固和迁移的阶段。这一步要做的第一件事是重新引入最初从任务中删去的复杂因素，注意，为了把心理一需要降到最低，每次只引入一种复杂因素。根据这种情况，我们给儿童下列类型的问题：



当再次引入每一种复杂因素时，人们能回过头来追忆最初所经历的一系列步骤。首先提醒儿童想起等号的意义；然后想起加号的意义。还要进行指导，特别是关于先要获得所缺加数所用的计算策略的细节方面。最后，要提供许多练习项目，以巩固儿童在每一阶段中所用的策略。这些练习项目包括各种不同的数目和格式(如，空格在右边，空格在左边)。

虽然这种分析只是作为一个例证提出来的，但是基于这种分析制订的学校课程，实际上已经由阿伦(Allan)戈尔德和我自己设计出来了。当对正规幼儿园的儿童实施这种课程时，其结果和先前报告过的皮亚杰任务的结果十分相似。在接受新皮亚杰任务

训练的幼儿园儿童中，有80%达到了后测标准。而控制组的儿童在同样长的时间内接受了标准的加利福尼亚算术辅助练习册(《现代学校数学：结构和使用》)其中有10%达到了标准(凯斯，1978a；戈尔德，1974)。后来，用与正规幼儿园儿童心理年龄相同的迟钝儿童做实验也获得了类似的结果。虽然还没有打算去准确找出达种效果真正的所在，但总的说来，有一点是清楚的，教学方法作为一个整体在课堂环境中能象在发展实验室里一样有效。

D.新皮亚杰学派研究方法的局限性与得失

虽然缺少加数的问题是课堂发展任务的一个良好例证，但它决不是唯一的例证。年幼和迟缓的学习者在计算周长，说出时间、分数、加法、读词以及解比等多种多样的问题当中，都显示出同样过于简化和导致错误的策略。在能体现出这些策略的每个领域中，初步研究的数据表明：新皮亚杰学派的研究方法在设计更有效的教学方而是很成功的〔弗雷译(Frazer)，1977；亨特，1977；斯坦倍克(Steinback)，1977，史蒂文斯(Stevens)，1977〕。然而必须指出，尽管上述这组任务种类繁多，但上述任务都具有两个突出的特点：(1)它们都要求获得复杂的认知技能，而不只是要求获得一个概念或者一组事实；(2)对这些任务来说，明确规定从一开始就要清楚阐明它们所要求的技能的性质。对于不属于这种情况的领域或任务，新皮亚杰学派的这种研究方法目前还不适用。①

① 值得注意的是，被认可适于运用策略技能的一些领域正在迅速扩大。例如：〔科林斯和比格斯(Biggs，1976)，吸取了哈勒姆(Halam，1967)和里斯(Rhys，1972)早期所作的研究，制定了在地理课和历史课中所运用的某些执行策略，伯雷特(Bereiter)，1977〕和斯卡达马利亚(1977a)在英语写作方面也进行了类似的分析，在我自己的研究中，我曾经打算对社会科学和批判性思维方面作这种分析(凯斯，1972b)。当这些方面的任务变得更加确定时，当有效的解答策略的性质变得更好理解时，那么，运用新皮亚杰学派的方法的可能性就变得更现实了。

尽管把讨论只限制在用新皮亚杰方法进行教学设计的那些任务领域内，但是，仍然提出了在各种教学目标中应该采纳哪一个的问题。要记住，很可能是：(1)试图通过前面部分描述的那套程序使儿童达到熟练完成作业者的运算水平；(2)试图教儿童一种较为简单的但对解决有关任务还是有效的运算程序，正如缺少加数的问题所证实的那样；或者(3)只选择给予儿童的任务类型，使他们不需要运用比他们的指令系统中现有的更为复杂的运算程序。

对许多任务来说，上述三个目标中的第三个目标是最实际的和最经济的。对于企图加快获得发展能力这种做法所提出的批评是得不偿失的，因为所要教的这些能力严格说来是学生在几年中不管怎样都能自发学到手的(凯斯，1973b)。这一批评也适用于在前节详细列出的那些技术和原则，这样说可能是合理的。尽管这些技术在加速获得象解决缺少加数问题那种技能是有益的，但是，对当前这些方法不做任何改变，只是把任务放在几年以后的课程中去，也能达到相同的效果。在某种意义上说，我们列举的这些技术，可以想象为类似于在一个狭隘拥挤得象只长颈瓶的港口里重新改建一条船那样的工艺；假如人们能找出一种办法绕过瓶颈般的隘道所造成的限制，假如人们能考虑到已经存在这种结构，那么人们所建立的新结构的优美和复杂性是没有限度的。但是，在什么条件下值得特意采用所列出的全部程序并试图在这种限制内工作呢？肯定不是在两年以后这条瓶颈般的隘道逐渐变得宽阔的时候；也不是在原先的结构在那段时间里自己变得同人们要建立的那种结构比较接近的时候。即使存在这种情况，这也似乎是：人们应当把这些任务延期到一个班的所有儿童有足够的能力来处理这些任务时才教。如果做到了这一点，那么一时的小小疏忽或疲劳就不会起如此重要的作用，教学的进度将是比较快的，儿童也将感到具有更大的优势和自信(罗沃，1973)。

然而，至少在下列四种情况中，这种推论是不适用的，而已提出的那些技术可能会应用得格外好。第一种情况是，所教的学生是智力落后儿童，顾名思义，智力落后者是指他的发展落后于他的同龄人，而他所达到的最终水平也是比较低的。事实上，正如韦克斯勒(1958年)所指出的那样，智力落后者最突出的一个特点，准确地说就是他们的信息加工能量测验的作业是很差的。如果一个个体十几岁时的能量只是2或3，那么很清楚，一味地等待他去获得较大容量的策略是不会得出什么结果的。此外，另一点也是清楚的，如果不能发现一种克服他们信息加工限制的技术，那么他就不能获得许多有用技能。

第二种情况是，我们曾经描述的那些方法可能对不容易掌握的材料是有价值的，即使这种材料的教学推迟到成年期进行，也是如此。成人的心理一功力几乎没有年幼儿童那样低，然而仍然会出现这种情况：他们会把一个简单的但以前是成功的策略在不适用的场合下误认为是适用的；也还有这种情况，学习一个新概念或新技能的要求大大地超过了他们的信息加工的能量负荷〔凯斯，1975；卡尼曼和特沃斯基(Tversky)，1973；斯梅兹伦德(Smedslund)，1967；沃森，1971〕。于是，在任何情况下，只要成人和智力落后者基本上处于同样的水平：也就是说他们的能量决不会比目前所具有的能量更大，所以只是推迟教学这种抉择并不是有效的。因此，当材料的复杂性超过正常成人力所能及的范围，就引出了所说的那种不正确的策略，已经提出的那些方法将再一次显示出它们是很有用的。

第三种情况是，这些方法可能适用于来自特别地区或少数民族地区的、作业落后于同龄人的儿童。不管落后的原因是什么，从社会的角度来看，结果很可能是十分相似的：由于个体不能掌握同龄人看来是十分容易获得的材料而感到自卑和不安。在这种情况

下，根据学习层次进行补救可能是很有益的〔加涅，1970年；哈切特(Hackett)，1971年；雷斯尼克，1967年〕。但是目前所用的方法可能会收到更多的益处。

第四种情况是，已提出的那些方法可能对所说的体现发展性的任务是有用的。也正是在这种情况下，由于这样或那样的理由把教学推迟几年的这种抉择是不能接受的。可能是这种任务必须要推迟到学生再没有兴趣学习这种材料的那个发展阶段上(如，青年期)；可能是有关任务是其它某种文化上有价值的技能的先决条件，推迟这种任务的教学就意味着要增加正规教育的年限；可能是及早获得有关技能就文化本身来说是如此的有价值，以致不能容忍作任何的推迟。从上述任何一种情况看，新皮亚杰学派的研究方法在诊断目前学习困难的原因并提出克服这些困难的教学程序上是很有价值的。

Ⅷ　结　论

看来，通过回顾所提出的一系列基本论点并把它置于一个更广阔的前景中来作出结论是很有意义的。教学发展理论的目的是为儿童最佳地获得有文化价值的技能和概念提供有效的基础。皮亚杰理论为建立教学发展理论提供了一个良好的开端，因为皮亚杰理论探讨的重点在于智力运算在学习过程中的重要性；在于这些智力运算具有高度结构的事实；在于智力运算在发展过程中要经历一个固定的重组顺序的事实；以及在于不一致的知觉在激活智力运算的重组上能起重要作用的事实。另一方面，皮亚杰理论还不足以建立一门教学发展理论，因为它还有三个重要的问题没有解决：一是，如何识别出关于学校一般所教的各种技能的运算结构；二是，如何评估儿童现有的运算功能水平；三是，如何把儿童从

其现有功能水平提高到所要求的水平上，或是如何改变教学方法使儿童不改变现有功能水平也能掌握教材。为了回答上述这些问题，就必须有一种新的理论，它能为智力运算的获得和运用提供更为详细的过程模型，并且对影响这个过程的环境因素和成熟因素作出详细的阐述。

近来，符合上述要求的理论已经发展起来了。在这个新的理论中，是用塞蒙最早提出的执行策略集合这种新的形式来描述运算结构的。运算结构的发展被描述为一个策略重组的过程，这个过程或者是由于越来越多地了解到某一问题领域的复杂性所引起的，或者是由于看到了不能达到特定目标所引起的。所提出的这种策略重组的过程模型，注意到了四种不同因素的作用：被试的心理一功力的大小；被试的认知风格；被试已有经验的复杂性和知觉组织及其被试的情感倾向。前三种因素的研究表明：它们对运算结构的获得做出了主要的贡献；在特定情境它们之间的相互作用是有规律的并且是可以预测的。

假设这种新理论已经表现出某种实际效果，那么用它来回答皮亚杰理论还没有解决的问题似乎是合情合理的。本文所作出的回答如下：第一，与某一单科有关的执行策略，可以通过详述儿童必须学会执行的任务目标和熟练完成任务者达到这些目标一般所经历的一步步过程来确定；第二，儿童现有的执行功能水平可以根据给他们的同样任务并分析他们设法达到这一目标所经历的实际步骤来评价；第三，儿童现有执行策略和期望达到的策略之间的差距可以通过下列任一方法弥补：(1)向他们提出一项既可以显示他们现有策略的不合适性又可以提供一个正确策略模型的任务；(2)找出在性质上比成人通常使用的策略更简单的执行策略并进行这种策略的教学；(3)找出在性质上较简单的任务以使儿童用现有的策略就可以解决。但无论采用哪一种方法，必须注意把儿童必须照

应的信息或线索的鲜明性提高到最大限度，并且把所要求的心理协调的复杂性降到最小限度。

新皮亚杰学派对教学的研究在许多方面类似于其它的现代研究。在教学前应当分析任务结构的这个建议是加涅一度所提倡的(1968)。应当根据这种分析来评价儿童对教学的准备,这一建议也是由加涅提出来的，并且成为其它大多数教学分析法的一个组成部分。最后，任务作业分析应当自始至终采取一步一步的描述形式，这一建议也不完全是新的。这是由其他几个受现代认知心理学革命影响的学习理论家所倡导或者付诸实践的〔安德森和福斯特(Faust),1973年〕。但是，有两个建议既可以看出新皮亚杰学派的研究方法极其明显地不同于其他现代的研究方法，又可以看出这种研究方法极其明显地有其发展的起源。这两个建议是:(1)儿童现有的功能水平不能简单地从他已获得的熟练作业的一些成分来评估，而一定要把熟练作业看成它本身是有着独特的组织结构的;(2)处理儿童现有组织结构的最佳程序就是一种把皮亚杰的平衡模型的要素与帕斯克尔—莱昂的影响平衡过程的作业参数模型相结合的程序。

朱法良 译

四、整体心理学模型在教学技术中的作用

帕特里克·萨佩斯 伊丽莎白·麦肯 马里奥·赞诺蒂

I. 引　　言

世界上有关各种技能的学习和各级学校中各科教学的教育心理学研究已有很久的历史了。但是，关于教育心理学的传统研究大多集中在对个别学习过程的微观分析上。一些最早的例子是桑代克关于阅读和初等算术的研究。桑代克的不朽著作《教育心理学》(1913—1934)第二卷就是专门论述学习心理学的。他的科学目标在于说明如何能用他称之为准备律、练习律、效果律的三条基本规律来解释象算术这种基本课程的学习。虽然，桑代克当时说的话今天听起来有点陌生，但是准备律主要意思是说刺激和反应之间的联结对满足机体来说是很重要的。练习律实际上有两个部分。其一是使用律，这就是说，当在刺激和反应之间形成联结时，由于使用而使联结的强度增加了。其二是失用律，这就是说，当刺激和反应之间所形成的联结在一个时期内不去使用时，这种联结的强度就减弱了。今天我们可以把练习律说成是由于反复尝试而使反应强度增加，可把失用律说成是消退律，但是基本意思仍如同桑代克所说的那样。三条规律中最有名的是效果律，这就是说，如果在刺激和反应之间形成了联结，并且伴随着一个满足的状态，用今天的话说是给予阳性强化，那么这种联结就增强了；同样，如果给予阴性强化，那么联结就减弱了。

桑代克在后来的一部著作《算术心理学》(1922)中，企图把他的这些学习原则逐条用于算术练习的心理学中。他把练习的心理学目的看成是加强刺激和反应之间的联结。人们往往认为桑代克把无意的机械练习强调为算术的核心，但实际上，在他的著作中，他始终强调了学生理解算术演绎推理的重要。纵然他没有为理解这种演绎推理的概念提出鲜明的智力依据。

总而言之，上述这一番题外话不是重把读者引向桑代克理论的详细评论上，而是强调，桑代克在本世纪初开创了一个新的历史局面。其后相当长的一个时期内，心理学家们相继进行了关于基本技能心理学的研究，包括算术、语言和阅读心理学的研究。萨佩斯(Suppes)，特曼(Termon)和布雷恩(Brain)(1968)对本世纪三十年代之前的许多好的算术实验研究进行了非常全面的总结，而关于这些技能方面的文献数量极多，尤其是阅读心理学方面的文献多得无法在这里一一介绍了。关于阅读方面的最新参考文献是由卡罗尔(Carroll)和查尔(Chall)(1975)编著的。

本文的中心议题是想通过实例令人信服地说明从桑代克及其后继者所主张的关于具体项目学习的详细分析，发展到关于学生进步和课程组织的整体模型的意义。值得向往的是这种整体模型立足于心理学的原理上，而主要变化在于忽略了每个反应的细节，着眼于以往文献中没有特别阐述的学习和作业这两个方面。第一个着眼点是看平均作业，而不是注意个别项目的结构。第二个着眼点是看一段相当长的时期内的进步，而不是看经过几次练习或测验后的短时进步。

我们可以通过二项教学技术的应用来说明对整体模型的强调。把整体模型的应用与教学技术联系起来不是偶然的，因为特别是在我们能对量的分析进行必要控制的技术时代里，从而无论从概念角度还是从实验角度来说，都会使这种模型的应用变得更

有意义。这里我们不讨论这种整体模型为什么不能用于一般的课堂环境。我们没有兴趣展开这样一个消极的论题，而是积极地确定在使用无线电、电视和计算机这三种主要教学技术手段中的任何一种进行教学时能做些什么。

在我们详细举例说明之前，想对整体模型做一、二点较为一般的说明。按照前面所说的教育心理学的传统和实验心理学的传统，对许多研究者来说，难于接受忽略每个反应细节的整体模型。我们想强调一点，我们对这些细节仍然感兴趣。事实上，在萨佩斯、特曼和布罗恩(1968)以及萨佩斯和马宁斯塔(Marningstar)(1972)发表的有关计算机辅助教学的研究中，关于个体学习和作业的分析内容是很充实的。但是，在这种广泛分析的基础上来发展整体理论是很困难的。在本文中我们将根据一般的心理学原理来论述整体模型。

对于没有详细研究过这些问题的读者来说，或许感到奇怪，为什么在教育或心理学文献中很少讨论关于适于课程组织的总体原则是什么。我们来举两个例子。第一个例子是根据什么原则决定在一个学日中，某一门学科的时间比例以及在这门学科中用于某些概念或技能的时间比例。就拿五年级的算术课来说吧，多少时间用于问题解决，多少时间用于分数运算。关于这些相对时间分配的问题没有进行过什么合理的讨论。第二个例子，或许比第一个例子更重要，那就是我们对怎样来考虑学生完成一门课程的问题缺乏讨论。例如，评价二年级学生阅读的进步，四年级学生语言技巧的进步，或七年级学生初等数学的进步的正确方法是什么，这方面的讨论很少。不是说这方面已出现什么争论，实际情况是这方面的详细技术讨论几乎就没有，我们不是说根本没有这方面的文献，但与课程组织和教学的重要性相比，显然没有作为一个主要的论题。

我们刚刚举的例子既庞大又复杂，可能会发展成更大、更复杂的问题。关于如何解决这些问题，我们不妄求作一般的回答。本文提出了较为适中的目标，不做更多的侈谈。我们想比较详细地举两个整体模型应用的例子，概略地叙述一下这些整体模型如何用来系统地考虑课程和教学问题。

第一个例子出自于我们所进行的计算机辅助教学方面的研究，并进一步发展了萨佩斯、弗莱彻(Fletcher)和赞诺蒂(1976)所开创的关于学生轨道的研究。第二个例子出自于最近在发展中国家使用无线电教学技术的研究。我们吸取了由西尔(Searle)、弗兰德(Friend)和萨佩斯(1976)所叙述的在尼加拉瓜开办无线电数学课外自修项目广播教学的经验。在这个例子中，我们关心的是集体学习而不是个体学习的教学组织的最优化问题。第二个例子比第一个例子更直接强调了最佳化问题。在第一个例子之后，我们做了一些说明，以便说明怎样用学生轨道理论来考虑教学最优化问题。

所举的两例都涉及到了小学阶段的基本技能，但更重要的是要看到，这种只限于小学的课程和教学的做法是无关紧要的。事实上，在第一个例子中所开发的轨道模型已用于大学的计算机辅助教学课程中，换句话说，已用于大学生的逻辑学入门教学中，拉森(Larsen)、马科森(Markosion)和萨佩斯(1978)已报告了这方面的应用。显然，第二个例子集中讨论的最优化问题也能用于其他更多的学科上，而不只限于小学的课程。

Ⅱ. 三门计算机辅助教学课程的学生轨道

A. 模型和理论

在这部分，我们将对萨佩斯等人(1976)所开创的评价课程的

方法提出新的探讨。这种新探讨的基本思想是对一个学生通过一门课程的进展情况提出预报理论，并用这种理论作为控制手段来调整某一个学生在这门功课上所用的时间量，由此对每个学生等级位置的增进达到了统一的目标。于是我们避免了传统评价法根据前测和后测来评价一门新课程所应采取的等着看的态度。

这种预报评价需要个别进行，因为指望所有的学生都获得同样的增进是不现实的，即使允许每个学生花不同的时间也是不可能的。一个直观的想法是研究一个学生通过一门课程的轨道，这种方法类似于物理学中物体运动的路线或轨迹的研究。预报学生进步的一个重要特点是把课程的总体特点与某个学生特有的全部个体参数区分开。具体是如何区分的，看下去就清楚了。

我们用计算机辅助教学的等级位置做为因变量，用接受计算机辅助教学的时间作为自变量，直接检验了前一段时间所探讨的关于幂一函数理论(萨佩斯等人1976)的健全性。如上所述，这种理论是指向学生总体进展，而不考虑他的个别练习作业。由于使用了这样一种整体方法，就有可能提出关于信息加工的一般公理，由此推导出一个基本的随机微分方程，这个方程似乎是几门不同课程所特有的。推导出的方程如下：

$$y(t)=bt^{k}+c,$$

该方程中的参数b、c和k对每个学生来说可以分别估计出来，就总体来说估计参数k常常是很方便的，也是很有意思的。为了随后参考时用，在此我们对每个学生和每门课程给出了下面的基本方程：

$$GP_{ij}=bt_{ij}^{k}+c,$$

这里，

Gp_{ij} ＝在第j次观测时学生i的等级位置；

t_{ij} ＝在第j次观测时学生i接受计算机辅助教学的时间，

b、c、h＝每个学生的估计参数。

在下面我们将要扼要评论的一些公理中，我们说到了信息的新单位(pieces)。采用这种非技术性的信息单位是免得连累真正信息单位(bit)的概念。实际上，这些公理就很少是采用关于信息这种方式来表述的。这可能或多或少是整体理论的一个特点，但不是一个优点。而详细的微观分析理论就要更多地谈到在某一门课程中传递给学生的信息结构。

第一个公理是在一个教程中一个学生加工或抽取信息的速率。第二个公理是当输入一个新单位信息时，学生信息加工的平均速率有什么变化。第三个公理是关于新信息输入速率的基本假设。第四个公理是在一个教程中学生当前的位置与到当前为止所输入的信息的总量密切相关。第五个公理是在这个教程中学生进步速率的一个类似的假设。读者如果对这些公理的技术性描述和基本随机微分方程的推导不感兴趣就可略去不读，而跳到下一部分。

为了说明这些公理和留待后用，我们规定了如下的量：$y(t)$＝学生在这个教程中的位置，就眼下来说当$t=0$时，我们设$y=0$，以后需要时再变；$dy(t)/dt$＝通过这个教程的进步速率；(At)＝到时间t时，在这个教程中所输入信息的累积量；$dA(t)/dt$＝在这个教程中输入信息的速率；$S(t)$＝学生加工或抽取信息的速率。

公理1，一个学生加工或抽取信息的平均速率$S(t)$与在一个教程中输入信息的速率成正比，与到时间t时输入信息的总量成反比，也就是说$S(t)$与$[dA(t)/dt]$成正比，与$A(t)$成反比。

公理2，输入一个新单位信息时，学生加工信息的新的平均速率降低了，其降低量等于他的当前速率与其渐近速率之差的积，这是一个很小的时间间隔h，

$$s(t+h)=s(t)-[s(t)-s(\infty)]s(t).$$

公理3，对某一个学生来说在时间t时输入一个新单位信息的概率依赖于时间t和以前的信息输入。

公理4，在一个教程中一个学生所处的位置与到此为止在这个教程中所输入的全部信息成正比，也就是$y(t)$与(At)成正比。

公理5，在一个教程中一个学生的进步速率与在这个教程中输入信息的速率成正比，也就是$dy(t)/dt$与$dA(t)/dt$成正比。

现在我们再来看看基本随机微分方程的推导。从公理3看，新信息的输入是一个普阿松过程，我们说具有参数λ。而从公理2看，由于在一个小的时间间隔h内概率为λh，所以

$$s(t+h)=s(t)-[s(t)-s(\infty)]s(t);$$

由于概率为$o(h)$，输入了一个单位以上的信息，并由于概率为$1-\lambda h-o(h)$，所以，

$$s(t+h)=s(t),$$

根据方程(1)和(2)，并且设$S(\infty)=0$，这样一来一眼就看清楚了

$$\frac{s(t+h)-s(t)}{h}=-\lambda s(t)^2+\frac{o(h)}{h}.$$

因此，当$h\to0$时，我们就得到了下面这个微分方程：

$$ds(t)/dt=-\lambda s^2(t),$$

其解为

$$s(t)=\frac{1}{\lambda t+c_1}.$$

根据公理1

$$s(t)=\frac{k_1 dA(t)/dt}{A(t)}, \qquad k_1>0,$$

而根据公理4

$$\frac{dy(t)}{dt}=k_2\frac{dA(t)}{dt}, \qquad k_2>0,$$

根据公理5

$$\frac{dy(t)}{dt}=k_3\frac{dA(t)}{dt}, \qquad k_3>0,$$

由此，合并结果为

$$\frac{dy(t)/dt}{y(t)}=k_3\frac{k_4}{\lambda t+c_1}, \qquad k_4>0,$$

求上述方程的积分，得到

$$\ln y(t)=\frac{k}{\lambda}\ln(\lambda t+c_1)+\ln|b_1|,$$

因此

$$y(t)=b_1(\lambda t+c_1)k_4/\lambda.$$

这里，如果$t=0$　$y(t)=0$，那么$c_1=0$。假定这个学生在$t=0$时对这门课程已有一些知识c，我们就得到了最终的方程

$$y(t)=bt^k+c.$$

B.学生的数据与课程的说明

为了检验这个模型，我们使用了美国几所聋人学校的耳聋学生和听觉损伤学生的学习数据和印第安事务局在新墨西哥州的伊斯利塔皮波罗开设的公立学校的印第安儿童的学习数据。在这些聋人学校的数据中，我们选用了佛罗里达、达拉斯(包括福特沃恩的105名学生)、俄克拉何马、休斯敦、圣安东尼奥和奥斯汀这六个地区的数据。这个已获数据的研究项目旨在为学校的计算机辅助教学提供辅助性服务，特别是通过数据评价和教师、学监培训这种

形式来为学校服务。在福利局(美国的教育机关)和国家科学基金会的倡议下，斯坦福大学社会科学数学研究所通过开展这项研究而把计算机辅助教学推广到学校，这是一项持续三年(从1970年7月到1973年6月)的庞大项目。当该研究所中央计算机传送的计算机辅助教学在三年计划结束，课程告一段落时，伊斯利塔学校和13所积极参与该项目的聋人学校中的10所与加利福尼亚的波罗阿托计算机课程公司(CCC)就硬件设备与课程签订了合同。新的硬件设备是由安装在每个学校的微型计算机组成的，学生使用电传打字机终端与计算机对话。

所收集到的主要数据是语言技巧，数学和阅读这三门课程中前测和后测成绩的分数和等级位置分数，基本上是每二周检查一次各门课程的等级位置。在后边关于课程的部分还要谈到如何周密地组织课程以便提供这个等级位置。因为各学校学生接受计算机辅助教学的程度有相当大的变化性，所以我们在表4.1的P列给出了每个测位上接受计算机辅助教学的学生数，在I列给出了在某一

接受CAI教学的学生数(p)及进行区域和课程分析所包含的学生数(I)

表4.1

课程	佛		达		伊		俄		德		休		总数	
	P	I	P	I	P	I	P	I	P	I	P	I	P	I
数学	251	175	388	163	184	128	197	146	650	341	198	90	1868	1043
阅读			388	97	184	70	197	49	650	55	198	17	1617	288
语言			388	204	184	78	197	102	650	297	198	45	1617	726

注：佛—佛罗里达，达—达拉斯，伊—伊斯利塔，俄—俄克拉何马，德—德克萨斯，休—休斯敦。

门课程中至少有3/10年等级位置变化(上、下）的学生数，也就是说，通过一年的教程至少发生了3/10年的变化。由于种种原因，几乎每个测位都有许多学生半途而废，只接受了短时期的计算机辅助教学。因此，我们在分析时排除了他们。尽管如此，剩余的学生数仍然是很多的，为我们的模型的统计分析提供了足够的根据。

对该项目的数据进行分析是很复杂的。原因之一是各学校所使用的成绩测验很不相同。表4.2列举了各学校所使用的测验代号和相应的测验名称。表中所列的这些测验代号及其所代表的具体测验后边还要用到。

表4.2　成绩测验的代号

代号	含　　意
1	斯坦福成绩测验，初级水平1，A[a]式
2	斯坦福成绩测验，水平2，A[a]式
3	斯坦福成绩测验，高级组A[a]式
4	斯坦福成绩测验，初级1组，W式1964[b]
5	斯坦福成绩测验，初级2组，W式1964[b]
6	斯坦福成绩测验，中级2组，Y式1965[b]
7	斯坦福成绩测验，高级组，W式1964[b]
8	斯坦福成绩测验，中级1组，W式1964[b]
9	斯坦福成绩测验，中级2组，W式1964[b]
10	科学研究联合会，成绩系列，初级2，E式1972[c]
11	科学研究联合会，成绩系列，初级2，F式1972[c]
12	科学研究联合会，成绩系列，多级水平，EF式1972[c]

a　只用于听觉损伤学生作为特殊成绩测验程序的一部分，1972年由华盛顿市的听觉损伤儿童和青年年评、人口学研究机关，高尔劳德学院设计的，此外还有理查德·马登，埃里克F·加德纳，赫伯特C·鲁德曼，查克C·默温，纽约公司、哈考特·布雷斯·乔瓦诺维奇也参与了设计。

b　是特鲁曼L.凯利，理查德·马登，埃里克F·加德纳，赫伯特C·鲁德曼，哈考特·布雷斯·乔瓦诺维奇，纽约公司设计的。

C　是罗伯特A·内斯伦德，路易斯P·索普，D·韦尔蒂·莱弗，科学研究协会，芝加哥公司设计的。

我们取样的学生接受了计算机课程公司的一门或多门课程的教学:数学课,1—6年级;阅读课,3—6年级;语言技巧课,3—6年级。我们将扼要地谈一下每门课程。先从解释单元教学策略开始。若想进一步了解细节可参阅每门课程的教师手册〔阿德金斯(Adkins)和汉密尔顿(Hamilton)1975;弗莱彻、阿德金斯和汉密尔顿1972;萨佩斯、西尔、卡恩茨(Kanz)和克林顿(Clinton)1975〕。

1.单元策略

一个单元代表一门课程中的一方面的内容。例如，数学课包括了除法单元,小数单元和方程单元。每个单元又是由一系列有联系的项目组成的,其顺序由易到难。一个计算机程序分别记录下学生每个单元的位置和作业。通过学生完成一个单元教材的作业记录与预先规定的作业标准的比较程序就可以确定这个学生是否还需要在本单元内的同一难度水平上再多做练习，是否应当退到较容易的水平上进行补课，还是已经掌握了现有的概念，可以往下进行,进入到更难的水平。因此,程序可以把学生的位置自动地调整到这个单元的适当水平上。这种评价和调整过程可用于所有单元中，而且贯彻到每个学生学习每门课程的始终。

按教材内容的难度水平均匀地分等级以便使得一个单元中的不同位置可以与学校的一年分十等分的等级位置相匹配。学生某一具体科目方面的等级位置可通过检查学生在代表那个科目方面的单元中的位置来决定。因为每个单元的作业都被分别地记录下来，并加以评定，所以在一门课程的每个单元中，学生都可以有不同的等级位置。教师的报告也是每门课程教学的一个组成部分,学生的进步是通过教师在报告时指出学生在每个单元的等级位置而记录下来的。

在一门采用了单元教学策略的课程中，一节正常课是由不同

单元中的练习混合而成的。每次只要呈现出某一门课程的一个项目，计算机程序就随机选择出提供这个练习的那个单元。随机选择单元可以保证让学生得到各种不同类型项目的混合而不是一系列类似的项目。

当学生开始学习一门课程时，每门课程在所有单元中都提供了位置调整。那些在他所进入的等级水平上完成得很好的学生可以以半年一步的跨度上升，直到更高挑战性的水平时为止。完成得不好的学生就按半年一步的跨度下降。这种总的等级水平调整保证了学生可在这门课程中居于正确的位置上，实际上只是在前10个课时对一个学生进行了这种总的等级水平调整。

2.数　学

1—6 年级的数学课包括14个单元的内容。表4.3列举出了数学课的这14个单元的内容。课程从第一等级水平开始并延续到第7.9等级水平。第7等级的教材不能构成那个等级年度的完整课程，只打算为完成第6等级教材的学生补充一些知识。

把每个单元分成若干份量相等的课业，或若干套数量、性质和结构类似的练习。在每一计算机辅助教学的数学课时，学生所得到的练习来自每个单元，这些单元包括了相当于他们的等级水平的那些等量的课业。例如，一个处于2.0等级水平的学生就从表4.3的前七个单元中得到练习。学生不是从所有单元中得到同样数量的练习。计算机程序调整每个单元的练习比例，以便与一本普通教课书所涉及到的相应概念的练习比例相匹配。

1—6年级数学课教材没有贮存在计算机的记忆中，而是采用了随机数技术的算法形式给予练习。当某一等量课业被选择时，一个程序就生成了练习所用的数值，产生了呈示这个练习所需要的格式信息，并且计算出正确的反应，以便与学生的输入相比较。

表4.3　　　　　数学单元　(1—6年级)

单　元	内　　　容	缩　写
1	数概念	NC
2	横式加法	HA
3	横式减法	HS
4	竖式加法	VA
5	竖式减法	VS
6	方程式	EQ
7	度量	MS
8	横式乘法	HM
9	算术运算规律	LM
10	竖式乘法	VM
11	除法	DV
12	分数	FR
13	小数	DC
14	负数	NG

结果，课时安排和实际给的练习在水平相同的学生之间是不同的，在一个几门课程的等级位置保持不变的学生的各种课程之间也是不同的。

3.阅　读

3—6年级的阅读课是由旨在改进学生五个方面的技能的阅读练习项目组成的，这五个方面是字词分析、词汇量的扩大、句子结构的理解、书面材料的解释和学习技能的发展(见表4.4)。包括了三、四、五、六四个年级等级水平所要学的教材以及降到2.5等级水平的补课材料

这个程序分成两个部分：基本句子和单元。基本句子部分从2.5等级水平开始，到3.5等级水平结束。这个部分的项目是短小

表4.4　阅读单元(3—6年级)

单元	内　　容
A	分析字词——把字词分解成若干单位
B	词汇——建立起阅读词汇
C	词义理解——理解句子和短语中词的意义
D	解释性的理解——阅读句子以便解释
E	现场学习技能——学会有效地运用资源

而容易的。这个部分的项目是现代计算机辅助教学系统所能显示的最简单的阅读训练练习。而单元部分从3.5等级水平开始，并且持续到6.9等级水平，当学生从事这些单元的学习时，在每个课时都会得到来自所有五个单元的项目。

4.语言技巧

3—6年级的语言技巧课包括了3—6年级每个等级水平足够用的材料。还为具有特殊语言问题的学生提供了补充课。有特殊语言问题的学生是指那些听觉损伤的学生和英语为第二语言的学生。

语言技巧课强调用法而不是语法，很少出现语法术语。语言技巧课又分成两个部分，语言技巧单元和语言技巧专题。这两种课程涉及到了相同的学科领域，但是其结构是很不同的。语言技巧单元使用了单元结构，以便提供高度个体化的混合练习。而语言技巧专题是全班共同接受教师指定的同一个专题的课程。这次所调查的学校只使用了语言技巧单元课程。要知道语言技巧单元的内容，请看表4.5。

表4.5　　语言艺术单元　(3—6年级)

单元	内　　容
A	动词的主要变化形式
B	动词的习惯用法
C	主谓一致
D	代词的用法
E	省略、所有格、否定语
F	修饰语
G	句子结构
H	构成法

C.结　果

正如我们在前面所说的(见ⅡA),我们所要检验的宏观模型是

$$GPi_j = bt_{ij}^k + c,$$

这里

GPij＝在第j次观测时学生i的等级位置

tij＝在第j次观测时学生i接受计算机辅助教学的时间

b、c、k＝每个学生的估计参数

表4.6　当K值的平均平方离差降到最小时的学生数

K	DL	DM	DR	FM	HL	HM	HR	OL	OM	OR	TL	TM	TR	IL	IM	IR
0.001	0	3	13	4	0	1	0	0	1	4	1	6	6	0	1	6
0.002	56	8	16	8	6	7	0	22	1	3	3	20	1	20	3	7
0.003	2	13	8	9	0	8	1	0	5	0	1	21	2	0	13	1
0.004	1	8	2	16	0	7	0	0	7	0	0	23	1	0	8	1
0.005	0	0	1	11	0	4	0	0	7	2	0	16	4	0	9	1
0.006	0	1	0	12	0	2	0	0	6	0	0	12	1	0	4	1
0.007	1	0	0	10	0	0	0	0	7	2	0	7	1	0	9	1
0.010	0	2	0	2	0	0	0	0	8	0	0	10	0	1	4	2
0.020	1	4	0	1	0	0	1	0	2	1	0	3	1	0	2	2
0.030	2	3	0	2	0	0	0	0	0	0	0	0	1	0	0	0
0.040	1	0	0	0	0	1	0	0	1	0	0	0	1	0	0	1
0.050	0	1	0	0	0	0	0	1	2	0	0	0	0	0	0	0
0.060	0	0	0	0	0	0	0	0	1	2	0	1	1	0	0	1
0.070	0	0	0	1	0	2	0	0	0	0	0	1	0	0	0	0
0.080	1	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0
0.090	0	2	1	1	0	0	0	0	1	0	1	1	0	0	0	2
0.100	6	5	3	1	0	1	0	0	2	2	0	4	0	2	2	0
0.150	2	1	2	2	0	0	0	0	2	2	0	7	0	2	0	3
0.200	3	9	0	3	1	1	1	0	1	1	1	5	1	1	1	0
0.250	12	6	3	3	2	0	0	1	1	1	0	8	0	1	4	1
0.300	9	7	0	3	0	3	0	1	2	1	0	3	0	1	3	2
0.350	10	7	1	1	2	3	0	1	3	1	0	6	1	3	2	0
0.400	7	8	1	2	1	2	0	2	1	1	0	5	2	3	3	1
0.450	7	8	0	2	2	1	0	1	1	0	1	3	0	2	0	0
0.500	7	6	2	3	0	1	0	0	2	1	1	3	3	0	2	0
0.550	5	8	1	7	3	3	0	2	4	0	1	9	1	2	1	0
0.600	6	5	0	2	1	2	2	3	2	1	0	2	1	4	2	2
0.650	4	4	1	1	1	0	0	0	1	0	0	10	0	4	1	0
0.700	5	5	1	1	0	1	1	2	1	1	1	5	0	2	1	1
0.750	2	4	1	1	1	1	0	6	0	3	0	10	0	0	1	1

续表

K	DL	DM	DR	FM	HL	HM	HR	OL	OM	OR	TL	TM	TR	IL	IM	IR
0.800	3	4	1	4	0	1	0	3	3	2	0	10	0	4	1	0
0.850	2	3	1	1	0	0	2	2	1	0	0	8	0	2	3	0
0.900	4	4	2	1	2	2	1	4	0	0	0	4	0	5	3	1
0.950	4	2	3	3	0	1	1	3	1	0	1	5	4	1	1	0
1.00	1	0	3	3	0	2	1	0	1	0	1	5	0	2	0	0
1.05	5	1	3	0	0	5	0	3	2	1	1	3	0	2	3	0
1.10	1	1	0	4	2	1	1	4	3	1	0	8	1	0	1	1
1.15	1	5	1	0	0	3	1	2	1	0	0	9	0	3	1	1
1.20	1	2	0	0	0	2	0	1	3	0	0	2	2	1	3	1
1.25	0	3	1	5	2	2	0	1	3	0	1	3	0	1	4	2
1.30	4	1	3	4	0	2	0	3	5	0	9	12	0	1	0	2
1.40	1	0	1	2	2	0	2	5	4	3	8	6	2	0	4	2
1.50	6	3	3	2	2	3	1	0	3	1	6	2	0	0	2	3
1.60	1	0	2	5	0	0	0	1	2	1	4	9	0	0	1	0
1.70	1	0	3	3	1	1	0	2	3	0	9	6	0	1	2	2
1.80	2	0	1	2	1	1	0	0	3	1	9	2	0	1	4	0
1.90	0	0	1	3	2	1	0	3	3	1	5	3	1	1	1	1
2.00	5	2	3	12	1	7	0	5	7	2	114	14	3	0	3	3
3.00	8	2	3	6	5	2	0	8	11	0	102	16	1	1	3	5
4.00	4	2	5	6	5	2	1	10	15	7	16	12	12	3	12	9
总 数	204	163	97	175	45	90	17	102	146	49	297	341	55	78	128	70

注： 表头中的每个缩写字其中一个大写字母D，F，H，O，T，I分别代表达拉斯，佛罗里达，休斯敦，俄克拉何马，德克萨斯，伊斯利塔皮波罗地区；L，M，R分别代表数学，阅读和语言技巧课。

现就0.001—4.00之间的50个K值（表4.6的第一列给出了K的确切值）来估计每个学生的参数b和c。使标准误降到最小时的这三个参数(b，c，k)和每个学生的标准误都记录下来。

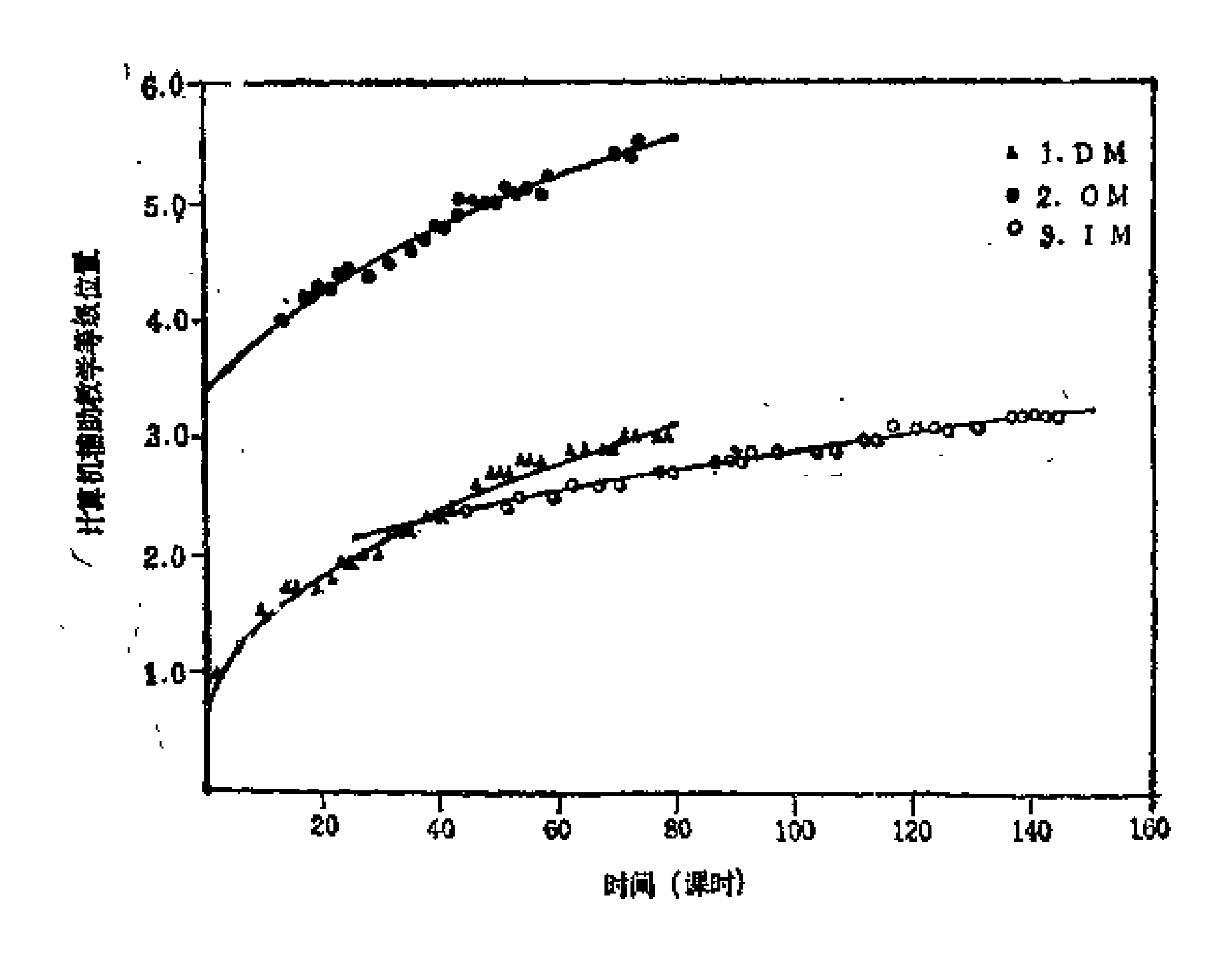


图4.1　在数学课上三个学生随时间的推移所获等级位置的散布区和理论轨道

图4.1到4.3表明这种整体模型多么拟合于各种学生的轨道，纵坐标是计算机辅助教学的等级位置，横坐标是接受计算机辅助教学的时间。图4.1表明了三个学生学习数学课①的数据。三个学生都稳步地前进，但是学生3比学生1和学生2都进步得慢。这三条曲线都与实测的数据点很贴近。学生1的数据符合于曲线

① 在该图和以后的图表中使用了如下的标志：第一个大写字母代表下述地区，D代表达拉斯，F代表佛罗里达，H代表休斯敦，I代表伊斯利塔皮波罗，O代表俄克拉何马，T代表德克萨斯（包括奥斯汀和圣安东尼奥的数据）。第二个大写字母M，R，L分别代表数学、阅读和语言技巧课。

$y=0.25\text{-}9t^{0.60}+0.509$，标准误差为[1]0.72：这个学生用了80个10分钟的学时获得了相当于二年的进步 。学 生 2 的 理论曲线方程是 $y=3.142t^{0.20}-2.313$， 标准误差为0.252。学生3的曲线是 $y=0.118t^{0.55}+1.419$，标准误差为0.030。

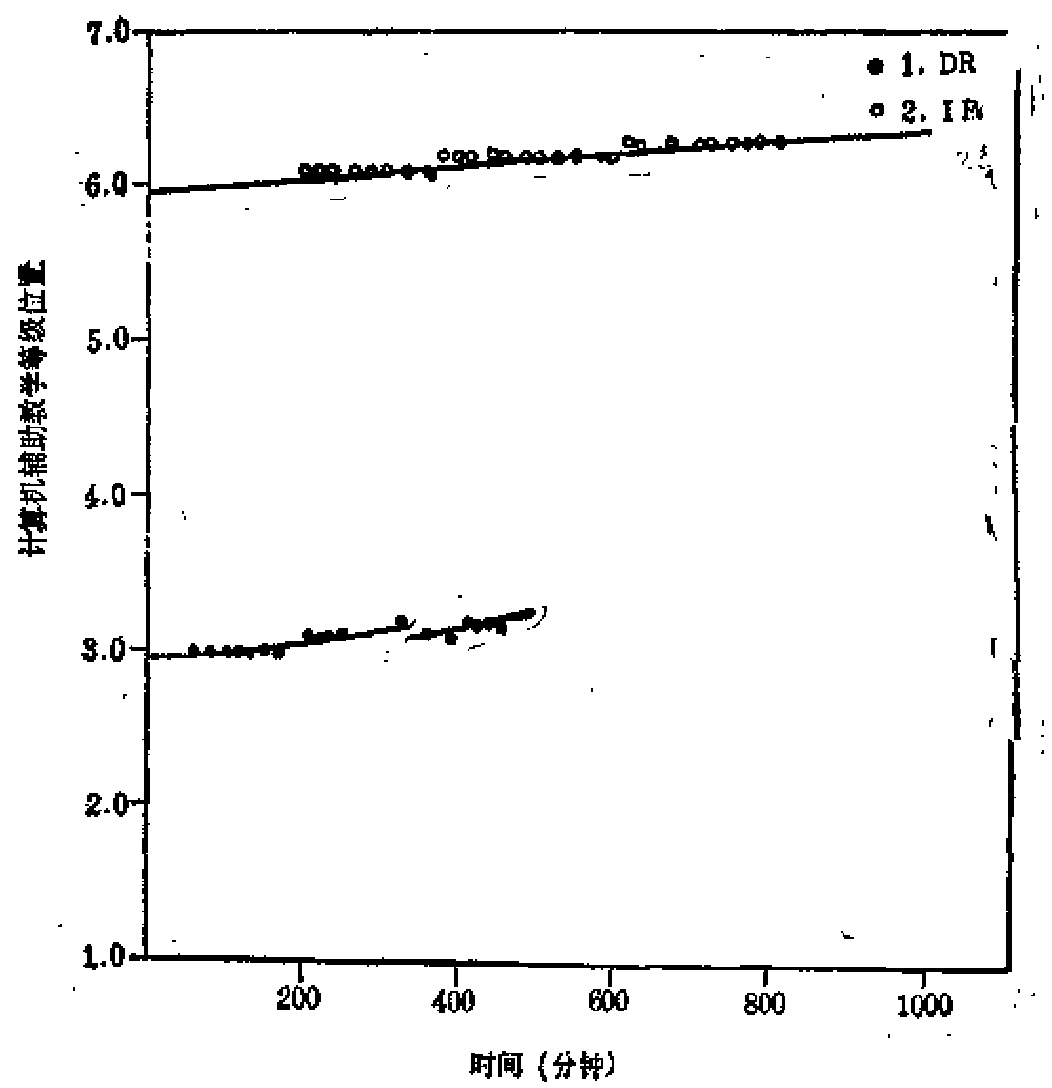


图4.2 在阅读课上两个学生随时间的推移所获等级位置的散布区和理论轨道

---

① 标准误是由下列方程限定的

$$SE_j=\left[\frac{1}{n_j}\sum_{i=1}^{n_j}(o_{ij}-t_{ij})^2\right]^{1/2}$$

这里，o_{ij}＝对学生j的观测i，t_{ij}＝相应的理论预报，n_j＝对学生j的预报数。因此学生样本总体的平均标准误差恰好就是这些学生的标准误差的平均数。

图4.2表明了两个学生阅读课的数据。请注意，横坐标上的时间单位是分钟。学生1用了50个10分钟的学时得到了1/3年的进步。学生2用了100个10分钟的学时得到了半年的进步。在一个学年中差不多有150天可以利用，每天都可以给学生安排一个学时计算机辅助教学阅读课。在此基础上，学生1，一个耳聋学生，用了1/3年的时间得到了1/3年的进步。学生2，一个美国印第安人，用了2/3年的时间得到了1/2年的进步，学生1的曲线是$y=0.00017t^{1.25}+2.935$。学生2的曲线是$y=0.009t^{0.8}+5.748$。标准误差分别是0.025和0.26

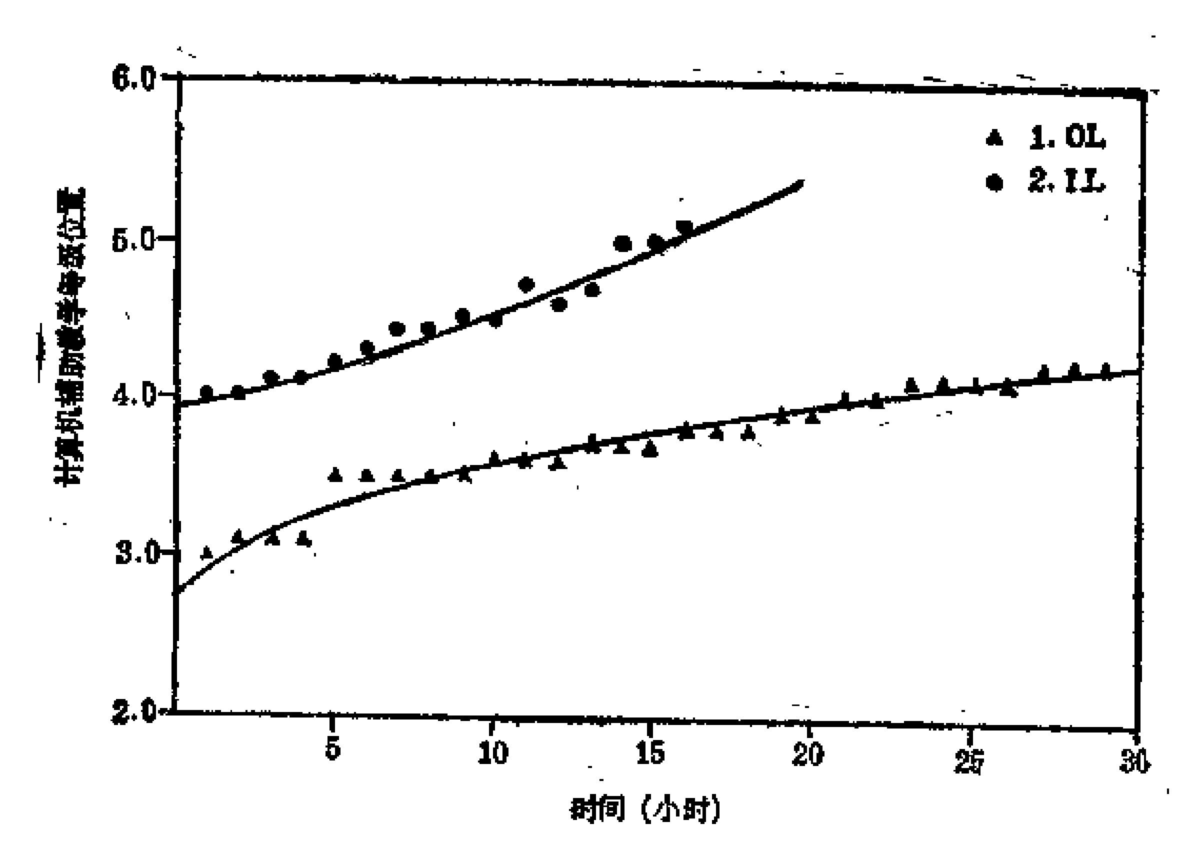


图4.3 在语言技巧课上两个学生随时间的推移所获等级位置的散布区和理论轨道

图4.3表明了两个学生语言技巧课的数据。两个学生都用了20个小时的时间而得到了一年以上的进步。学生2学了30个小时，得到了1.4年的进步。学生1的曲线是$y=0.035t^{1.25}+3.938$。学生2

的曲线是$y=0.224t^{5.5}+2.774$。标准误差分别是0.059和0.062。再一次看到，这些曲线都与实测数据相符程度是很好的。

把k的最大值的分布看作是各种不同学生进步速率的证据是有益的。表4.6表明了每组的K值最小标准误差的次数；地区和课程仍用前述的大写字母表示。

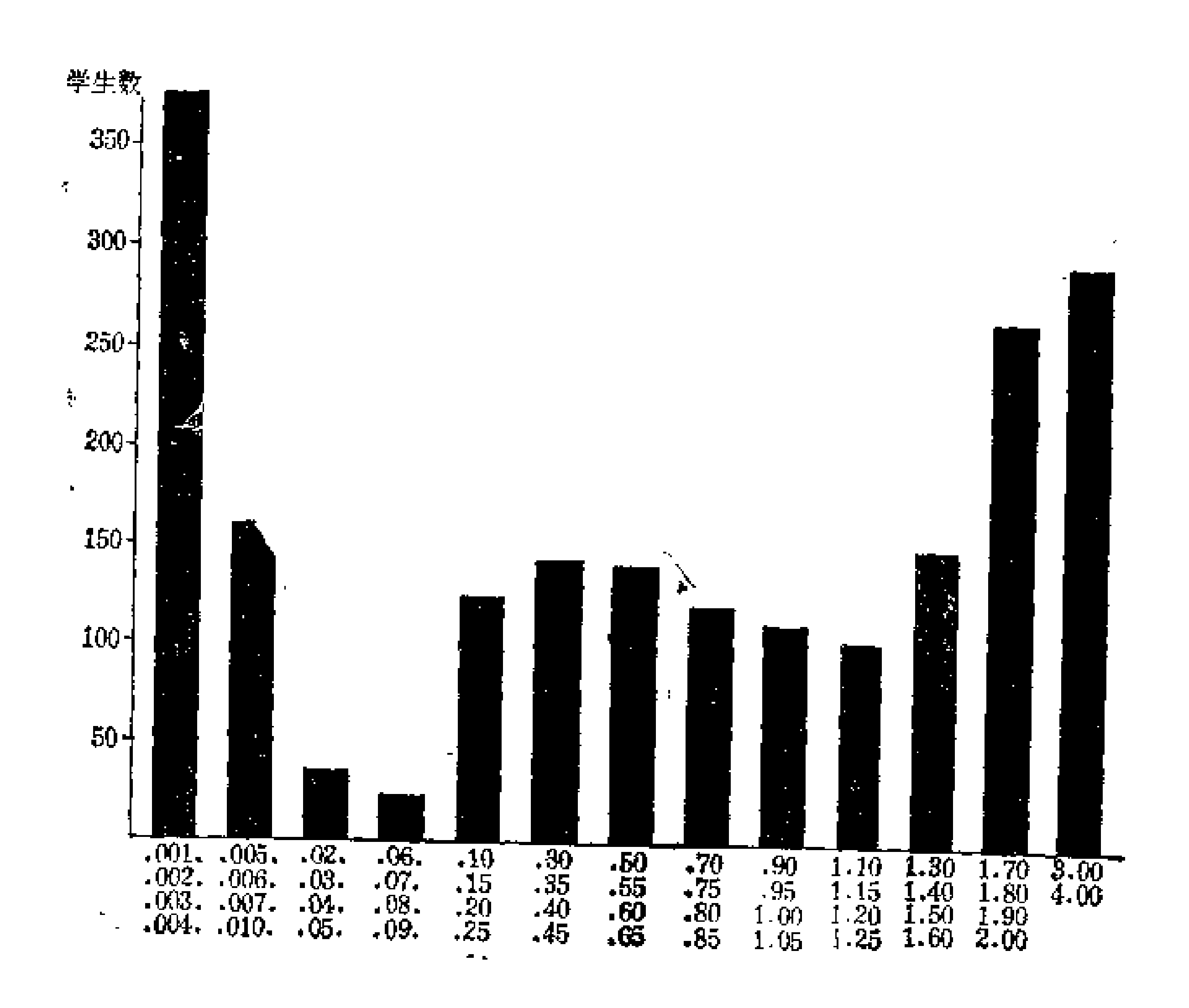


图4.4 对2057个学生的所有课程进行分别估计的指数k的直方图

图4.4是一个直方图，表明了对该项研究中所有被试学生逐个估计的K值的分布。对至少5个学生来说，每个K值都具有最小标准误。而对至少25个学生来说，大多数K值具有最小标准误。

在萨佩斯等人以前的著作中(1976)，总体固定K值平均标准误差的分布是比较平缓的。聋人学校和伊斯利塔的数据支持了这个结论。我们发现如果一个总体使用了固定的K值，那就不必要再进行精确地估计了。对于接受阅读和数学计算机辅助教学的大多数学生来说，在0.07—0.09之间的任何值都象其它估计值一样好，还可以适当调整另外两个个体参数b和c。例如，在伊斯利塔，数学课的这些K值的平均标准误大约在0.05—0.06之间。而对于语言技巧课来说，K值的最低界限是比较小的。事实上，接受语言技巧计算机辅助教学的四个地区中有三个地区是$K=0.002$的最小平均标准误。因此，在0.07—0.90之间的K值其标准误是不大的；例如，在俄克拉何马这些K值的平均标准误大约在0.2—0.3之间。

D.参数的联合分布

当我们估计每个学生的这几个参数时，很自然地就要问，参数的联合分布如何。其结果有二种情况，这要看对总体来说K值是否已固定，而对个体来说K值是否可选择。首先来看，为每个地区和每门课程所确定的一个总体K值，这是通过把逐一个体K值与每个学生因拟合$y=bit^k+c_i$方程而得到的逐对(b_1,c_{12})相关值加以平均而得到的。我们得到了16组这种相关值。除了三个组之外其余各组b和c的相关均接近于0。这三个例外是达拉斯的语言技巧课，其$r(b,c/k=0\cdot628)=-0\cdot6$，俄克拉何马的阅读课，其$r(b,c/k=1.081)=-0.9$，伊斯利塔的语言技巧课，其$r(b,c\cdot k=9.06=-0.8$。

另一方面，如果我们为每个学生找出这样三个参数(ki、bi、ci)，使其代入方程$y=bit^k+c$，能与每个人的实测数据最相拟合的话，那么这三个参数中只要有一个已定，其余二个参数之间的相关都是很高的。这可用数学式明确地表达出来，$|r(bi,ci)|\cong1$，

$|r(ki,ci)|\cong 1$, $|r(ki,bi)|\cong 1$。这就是说，若要与个体曲线拟合，三个变量中只有两个是必要的。就K值来说，结果表明，如果K值已定，那b和c就成为自变量，我们需要有这两个变量去与实测数据相拟合。但是，如果K值是因人而变的，那么b和c就成为因变量了。

对上述结果有几点不同的说明。第一，根据不充分的假设，关于条件独立性和任何两个参数之间具有绝对值为1的相关的结果，满足了萨佩斯和赞诺蒂(1976)所证明了概率引定理假设。换句话说，如果$r(bi,bi|k)=0$，并且$|r(bi,ci)|=1$，那么，$|r(bi,ki)|=|r(ci,ui)|=1$。第二，幂函数方程的幂级数表示指明了约束的来源。当$t>0$时，我们可以把t^k写成：

$$t^k=e^{k\ln t}=\sum_{n=0}^{\infty}\frac{k^n(\ln t)^n}{n!}$$

因此这个模型又可写成下列的同等形式。当$t>0$时

$$y=\sum_{n=0}^{\infty}a_n(\ln t)^n$$

其根据是：

$$a_0=b+c$$

$$a_n=\frac{bk^n}{n!}$$

特别是，就K的一个小值域来说，$b=a_1/K$，几乎是一条直线。关于这个数学模型这里不做更深的分析了。

1.预报后测分数

我们检验了使用参数b、c、k的个体估计值、最后等级位置(FGP)和前测分数(PRE)来预报出后测分数(POST)的可能性。这种预报数学模型可由下面这个方程来限定

$$\mathrm{POST}i=a_0+a_1(bi)+a_2(ci)+a_3(ki)+a_4(\mathrm{FGP}i)+a_5(\mathrm{PRE}i).$$

我们可用20组被试来检验这个模型。每组被试数 $N>6$,另外还要有前测和后测数据。其中有8个组$6<N\leqslant10$,有12个组$N>10$;这12个组中又有5个组$N>30$。在下面的讨论中,各组的名称由依次排列的四个项目(地区,课程,前测代号,后测代号)来表示。前测和后测代号的含义请看表4.2。

在有6—10个被试的那8个组中,除二组外,其余各组的多重相关均大于0.8(尽管这些小样本的多重相关很高,但我们并不看重它们)。两个例外是(俄克拉何马,阅读课,4,8,)和(德克萨斯,语言技巧课,6,1)。这二个组与所有20个组都很不一致,就(德克萨斯,语言技巧课,6,1,)来说,当所有的自变量都进入回归以后,在残数中只剩下四个自由度,多重相关还是0.562。这8个组中有5个组,在第一次或第二次重复时,参数b就进入了回归,在这个回归组中关于自变量的进入回归没有什么其他明显的倾向。

在有10个以上被试的那12个组中,有7个组的回归其多重相关大于0.85;(俄克拉何马,数学课,4,9)其多重相关是 0.967,自由度是5;(俄克拉何马,数学课,4,5)其多重相关是0.941,自由度是8;(俄克拉何马,数学课,5,9)其多重相关是0.914,自由度是13;(德克萨斯,数学课,4,4)其多重相关是0.862,自由度是20;(伊斯利塔,数学课,10,11),这是最大的一组,其多重相关是0.864,自由度是64。这12个比较大的组中,有8个组其参数b在最后或几乎最后才

进入回归，这与N≤10的那8个组的结果形成了明显的对照。对这，些较大的组来说，FGP和PRE是重要的自变量。12个组中有7个组FGP是进入回归的第一个自变量。12个组中有8个组，PRE是第一个或第二个进入回归的自变量。在这些回归中，一个不太重要的自变量的倾向是K，它在第一次重复时从未进入过回归，在第二次重复时进入回归也仅是一次。

虽然前测往往是一个很重要的自变量，但是有许多学生前测数据对他们是不可用的，这样就不能用上述的数学模型了。除此而外有一个不需要前测数据作自变量的数学模型允许地区相同，课程相同，后测归组了的那些学生使用。于是，就考虑使用下面这个模型以便增加自由度。

$$\mathrm{POST}i=a_0+a_1(bi)+a_2(ci)+a_3(ki)+a_4(\mathrm{FGP}i) \qquad (3)$$

我们能用N>5的29组被试数据检验这个模型。在下面的讨论中，各组的名称用依次排列的三个项目(地区，课程，后测代号)来表示。

在这些组的回归中，其中有几个例子用自变量的线性函数显然不能充分地解释实测数据。一个极端的例子是(德克萨斯，数学课，1)其计算结果是多重相关0.190，自由度48。但是，被试数达11或11个以上的24个组中，有8个组的回归其多重相关在0.8以上。学生数达25或25个以上的16个组中，有8个组的回归，其多重相关在0.7以上。29个组中有17个组，FGP首先进入了逐步回归的推算。

为了进一步扩大被试组，我们检验了方程(3)根据课程和后测跨地区归组学生的模型。这样一来，对这种归组来说有9个回归是新的(归组增加了抽样人数的那些回归)。其中有7个组被试人数超过10人，有6个组被试人数超过30人。在讨论这些回归时，各组的各称用依次排列的两个项目(课程，后测代号)表示。

在这些组中，有几个组的多重相关是很高的(语言技巧课,5)。其多重相关是 0.816，自由度是21。FGP 首先进入回归，然后是c和k，b没有进入回归，因为b和c之间有很高的负相关，$r(b,c)=-0.999$(语言技巧课,8)。其多重相关是0.812，自由度是29，自变量进入回归的顺序仍然是FGP、c，k，而没有b，因为$r(b,c)=-0.995$(数学课,5)。其多重相关是0.853，自由度是36(数学课,4)。其多重相关是0.897，自由度是35(阅读课,5)。其多重相关是0.753，自由度是3。11个组中除一组例外，其余各组FGP均首先进入回归。

E.关于最佳化问题的几点结论及应用

用幂函数模型表示的整体轨道模型在大量的例子中是与学生进步的个体数据非常吻合的。这种与数据的紧密适合支持了刚刚报告过的关于这个模型适合性的广泛分析（萨佩斯、弗莱彻、赞诺蒂1975,1976）。我们发现语言技巧课、数学课和阅读课的计算机辅助教学成绩与学年终了的考试成绩紧密相关。在数学课中得到了最高的相关，根据听觉损伤的学生掌握阅读或语言技巧技能会比数学更困难，可以预料到这一点。

轨道模型非常成功地适合于个体实测数据这一事实为教学技术应用最佳化指明了一些可能的方面。在此，我们只扼要地谈几个问题，并作较详细的分析，随后将发表。第一件要注意的事是根据花在系统练习、实习和复习上的时间量，一个轨道模型有可能做出学生等级位置的整体预报。考虑到对学校使用计算机的费用必须进行一定的分配，自然提出了这样的问题，分配给每个学生的机时应该是多少。这样一个机时分配问题接着又导致了另一个问题，我们设法达到最佳化的是什么功能。下面我们提出几种可能的选择。

第一种可能性是，最大限度地提高预期的等级位置，正如通

过课程等级位置或该教学之外的前测和后测所测量的那样。如果我们对可采用计算机辅助教学的已知总体指望达到这样一个无限的目标，那么其结果必然是把那些空着的或零碎时间分配给每个单元时间内等级位置提高得最快的学生。一般说来，这将意味着分配给比较聪明和比较好的学生的时间与分配给差等生的时间相反。第二种可能性是采取相反的办法——也就是说，我们的目标不是最大限度地提高所定总体的预期等级位置，而是缩小该总体等级位置的方差。在这种情况下，我们的最佳化的功能就是缩小最终等级位置或等级位置增益的方差。上述二种方案，无论选择哪种，其定性结果都是把大部分时间分配给差生，而对能力最强的学生基本上没给什么时间。（我们是用定性的而且是较笼统的词来描述上述这些最佳化方案的，因此就能很容易地看出如何把定性结果变成精确的定量公式了。）我们认为上述这二种极端的作法确实都不符合技术资源的实际分配。

第三种时间分配的形式是使预期的增益最佳化，但有一个限制，总体的等级水平成绩的方差不增加。第四种可能性是最大限度地扩大在等级位置上获至少一年增益的学生数。从多方面看来，只有最后一种目标更符合许多教育界人士的直觉的想法。

重要的问题是要认识到，上述四种最佳化的目标各自有一个不同于一般的机会均等哲学的特殊分配算法。也许我们想要对机会采取平均主义的态度，坚持把时间平均分配给总体中的所有学生。但是我们怀疑机会均等哲学的一些方面一定要运用于我们所讨论的任何整体模型的时间分配上。这就导致了关于最佳化的一个公式表达，在工作精神上可能接近于许多人的直觉思维。例如，假定进行了这种分配以后，还有未用过的技术手段，这个模型就以此设想为根据首先采取了机会均等的原则，并把每日的常规时间分给每个学生，然后分配剩余的手段，以便使我们上而找到的

功能中的一种最佳化。许多人可能选择了通过缩小方差来进行分配，重要的是要认识到，根据机会均等原则作出分配方案后，再使用缩小方差的整体模型与用缩小方差作为进行分配的唯一目标完全不是一码事。

在这一部分列举了我们所开发的这种整体模型可能应用的范围，但是决不限于上述大致介绍的几个方面。关于最佳化问题有一种不同的，从各方面看来可能是更有吸引力的考虑。不是迫使某些外部功能最佳化，而是运用这种轨道模型的预报作用与个别学生一起工作达到师生共同商定的一个目标。例如，一个学生可以和他的教师一道努力去实现1.1年的等级位置增益。可用轨道模型向教师，向奔向目标的学生连续提供最新的预报，并根据预报把时间分配给那些需要的学生。对我们来说，很容易低估年幼儿童能注意关于自己进步的信息并采取相应行动这一情况。从我们最近所做的预备研究来看，我们发现确实可以激励学生去达到有关课程的目标，并且根据他们自己的课程表找出利用学校中可用的额外资源的途径和办法。情况可能是这样，根据个人的目标及其主动精神来分配时间，并利用整体模型有效地进行预报和指导，这将会提供一种获得进步的较好方式。我们可以避开那种不考虑学生能进行自我激励的资源分配。

Ⅲ.数学广播教学课程设计的最佳化

使用无线电广播进行教学已有很长的历史了。但是，有关无线电广播教学技术的发展和改进方面的研究很少。目前世界上所采用的各种教学技术，其中包括程序教学手册、彩色电视和计算机，要以无线电广播教学为最佳，它对于大面积教育来说是最经济的。应发展中国家广播教学日益增长的应用和研究的需要，美

国国际开发署教育和人类资源处的技术援助局为改进小学低年级广播教学的效果提供了研究经费。在美国国际开发署的支持下，社会科学数学研究所发起了一个小学低年级数学广播教学的研究计划，目前正在尼加拉瓜实施

数学广播教学项目要对所有接受数学教学的儿童负责任。每天的课程是先收听25分钟的数学广播教学，然后进行大约20分钟的教师指导活动。教师的教学是根据项目预定的指导计划进行的。没有教科书，但是教师可以补充一些简单的材料，象瓶塞和小木棍之类。我们研究的中心问题是如何更好地使用这些有限的物力来进行有效的数学教学。

在讨论这个项目时，我们已发现有些人对数学广播教学持有怀疑态度，因此，我们这个研究项目的主要目的就是来验证我们所提出的假设。这项实验研究之所以成功，很重要的一点是由于广播教学的费用较低。如果教学效果良好，那么至少在小学数学教学中，广播教学在任何可采用的技术中具有费用最低而效益高这一优点。在费用低而效益高这一点上，甚至具有相应训练的教师所进行的传统教学也不及广播教学。

A.实施该项目的地点

这个研究项目定在尼加拉瓜的首都马那瓜附近的马萨亚行政区进行。这个行政区包括一个大城市和九个小城市。全行政区的小学生近18000人，其中三分之二的学生是城市学校的，其余的是乡村学校的。每班有15人至60人。乡村学校的一间教室里常常有几个年级的学生。大致说来，城市学校百分之三十的学生、乡村学校百分之五十的学生是一年级。

每年2月中旬至11月中旬为一学年，每天儿童上学3—4小时。在每个学年末由教师或学校出题进行期末考试来评定学生的

成绩。一门或二门课不及格的学生，在下一学年开始之前要进行补考，有二门以上不及格的学生就要留级。这个行政区的大多数教师是学校毕业生(具有五年小学后教育)。一般说来，这个行政区的教师水平比远离首都的边远地区的教师质量要高。到撰写本文的时候为止，已有几千名学生接受了广播教学，并且人数还在不断增加。

B.通观一节数学课

每节数学课是由广播讲座和教师指导这两个部分组成的，并且首先进行广播。就一节具有代表性的课来说，两个主持人与一两个协助者一道唱歌、游戏和谈论数学，并且常常邀请收听广播的儿童参加这些活动。在每25分钟的课时里，让儿童以各种不同的形式对50个练习做出回答。有时让儿童摆弄一些实物，如让他们数瓶塞或把瓶塞分成堆，有时让儿童对新的数学材料做出口头反应，有时儿童通过在自己的笔记本上写出练习和答案来进行个别反应。广播讲座结束后，教师接着上课，按照预定的教师指导计划进行。

C.课程结构

广义地说，数学课程是由尼加拉瓜教育部规定的，虽然在编排广播课时，在重点上做了某些必要的变化，并进行了某些必要的重组。根据教育部的一般课程指导来编排广播课有以下几个步骤。首先，把所要教的数学课的内容分成若干专题或单元，类似前一部分所说的计算机辅助教学的课程。就一年级的数学课来说，分6个单元，具体说是：基本概念，数概念，加法，减法，运用与度量。其次，就每个单元来说，要系统阐明一系列教学目标和任务，这就规定了成功地完成一年教学大纲的学生应具有的水平。这些

目标再细分成若干子目标，称为课，这就刚好与教学单元相符。①例如，我们来看一个目标，它是这样说的：学生将数出一组物体中的物体个数。这组物体是由N个物体组成的，这里N小于或等于25。第一个子目标可以把物体个数限制5个或少些，第二个子目标可以是6至10个物体，等等。这些课可以随时抽出来以便提供测验项目，测量学生获得课程目标的进步。

确定教学顺序的第一个考虑是，凡是做为某一课的先决条件的那些课要依次先进行，在一个单元内或各个单元之间也必须是这样。第二，在全年每一课必须出现几次——第一次出现是引进和解释材料，以后再出现是提供练习和复习。

D.达到不同课程目标的最佳时间分配

许多人承认，课程设计者和作者应该把教材所要达到的目标说清楚。一般说来，在数学广播教学的设计中，我们是力求这样做的。更重要的是，上述的课程结构自然便于定量研究。我们所提出的这种模型提供了特殊的定量结果，可以对那些编排课程的人提供充分的指导。

如上所述，在每节25分钟的广播课里，有差不多50个练习。而每一学年有近150节课，这就是说，大约有7500个练习分配给各

① 在表4.7和4.8中，是用缩写字的形式来表示这些课程或子目标的。现在按字母顺序来说明这些缩写字：HA=horizoutal addition横式加法，HM=horizontal multiplication横式乘法，HS=horizontal subtraction横式减法，NC=Number Concepts，数概念，OA=oral addition，口头加法，OC=Oral combination S of addtiouanol subtruction，口头加减法混合运算，OM=Orul multiplication 口头乘法，OS=Oral aubstraction口头减法，SN=Seleeting numerals选择数字，SP=Substraction word Problems(oral) 减法文字题（口头），VA=Vertical addition 竖式加法，VS=Vertical Subtraction竖式减法，WN=Writing numeruls 书写数字。表中缩写字后边的数字是课数。

种目标。而最佳化问题就是把全部练习分成由N个练习组成的若干组，然后确定具有不同目标的各组练习的个数，用这样一种方法来保证做完一组练习后能达到最佳的作业。

可以想象，这个复杂的问题在可能的范围内有其他约束条件。我们的最佳化问题，有以下几个成分。第一要有目标个数s。第二要有指定给每个目标i的加权相关系数ri，并具有明显的约束条件$ri \geq 0$，并且$\Sigma ri=1$。第三要有对某一组练习中的一个练习做出正确反应的初始概率，这个正确反应的初始概率我们用p_i表示，当需要表达出位置顺序j时，偶尔用p_{ij}来表示。与此同时，我们使用了一个表示错误概率的标准符号q_i和q_{ij}，并且$q_i=1-p_i$。第四要有平均学习速率a_i。请注意，不引入关于不同目标的项目的学习速率，就能使问题简化，但是，在此，我们没有做这样的简化。第五要有平均遗忘速率b_i。第六要有某一个组或片断的练习个数N，对第i个目标分配给m_i。例如在学年初，N是7500个练习。我们的任务是给每个目标分配m_i个练习，以便完成这组练习后能达到最佳效果。最佳化的衡量是错误平均概率期望值，因此，我们希望把一个错误平均概率期望值降到最低。

当然，针对不同的目标会有许多不同的方法呈示练习，但是我们决定先用周期轮换法。在涉及其它目标的k_i次试学之前的某一次试学中提供关于目标i的一个练习。在整个这组练习中k_i+1这种模式周期地反复出现。正如我们后边将指明的那样，显然k_i，m_i和N之间有一种关系，因此，k_i的值由m_i值和N值所决定。为了找出每个目标的练习数m_i的最佳分配，我们还要确定k_i值，而ki值是关于目标j的一个练习在下一次出现前的最佳试学次数。显然，在实际编写的教材中，有时会与$ki+1$这一严格的周期有所出入，但是，有一点小出入影响不大。而且由于使用了这种周期论换法，我们就可进行明确的计算，否则将是很困难的。

我们假设了所有学生的平均学习速率和遗忘速率。在这些平均假定中各种个体模型都能被拟合，正如从数学学习理论文献所得知的那样。特别是，符合于平均学习假定的个体学习模型一方面可以是全或无模型，另一方面也可以是线性增量模型。这两个平均方程如下：

学习： $q_{i,n+1}=a_iq_{i,n}$ (4)

遗忘： $p_{i,n+1}=b_ip_{i,n}$ (5)

根据错误概率我们可以写出遗忘的循环，

$$q_{i,n+1}=b_iq_{i,n}+(1-b_i) \tag{6}$$

在由一次学习性试学和k_i次遗忘性试学组成的一个k_i+1次试学循环后，我们能很容易地写出下列方程，

$$q_{i,ki+2}=b_i^{ki}a_iq_{i,1}+(1-b_i^{ki}) \tag{7}$$

萨佩斯(1964)找出了这个结果的求导，接着又找出另一些结果的求导。

在k_i+1次试学的m_i个循环以后，我们能给出下列方程

$$E(q_{i,m_ik_i+m_i+1})=b_i^{m_ik_i}a_i^{m_i}q_{i,1}$$

$$+(1-b_i^{m_ik_i}a_i^{m_i})\frac{1-b_i^{k_i}}{1-a_ib_i^{k_i}} \tag{8}$$

如前所示，N是在一个组内试学的总数。因为，根据假设，在所

有非学习性试学中发生了遗忘，所以在N，m_i，和k_i之间存在着如下简单关系

$$N = m_i k_i + m_i \qquad (9)$$

应当看到这个方程适用于每个目标i，当然具有如下的约束条件。

$$N = \sum m_i \qquad (10)$$

最后，为了得到错误概率q的加权期望值，在一组练习结束时，我们根据方程(8)，(9)、(10)得出下列方程：

$$E(q) = \sum r_i E(q_i, m_i k_i + m_i + 1) \qquad (11)$$

目标i的期望错误概率$E(q_i, m_i k_i + m_i + 1)$是$q_i$，1和$f(k_i) = (1 - b_i^{k_i})/(1 - a_i b_i^{k_i})$的一个凸复合（加权平均数）。如果要产生学习效果必须满足下列的不等式：

$$\frac{1 - b_i^{k_i}}{1 - a_i b_i^{k_i}} < q_{i,1} \qquad (12)$$

当出现这种情况时，$f(k_i)$是k_i的一个递增函数，我们就得到了下列式子：

$$\frac{1 - b_i}{1 - a_i b_i} \leqslant \frac{1 - b_i^{k_i}}{1 - a_i b_i^{k_i}} \leqslant E(q_i, m_i k_i + m_i + 1) \leqslant q_{i,1}. \qquad (13)$$

我们可以进一步得到$E(q)$的下介：

$$\Sigma r_i = \frac{1-b_i}{1-a_i b_i} \leqslant E(q). \qquad (14)$$

当(14)成立时，方程(11)就很难找出一个最优解——也就是说要把期望错误降到最低的那些学习性试学m_i的分配。有关详细分析结果将另文发表。

另一方面由于提出不存在遗忘这种十分简单的假设，也就是说$b_i=1$，所以我们能对分配问题得出一个明确的解，这使我们有了第一个近似值，在许多情况下都是很有用的近似值。因此，我们想把

$$E(q)=\Sigma r_i q_{i,1} a_i^{mi} \qquad (15)$$

降到最小，并且要附加$\Sigma mi=N$这个条件。我们使用拉格朗日乘数找出解(这里Z代替了标准入)。我们的基本函数式是

$$f(m_1,\cdots,m_s)=\Sigma r_i q_{i,1} a_i^{mi}+z(\Sigma mi-N).$$

于是进行求导，当$i=1,\cdots,s$时，我们得到如下式子，

$$r_i q_{i,1} a_i^{mi} \ln a_i + z = 0, \qquad (16)$$

并且

$$\Sigma m_i - N = 0,$$
$$m_i > 0.$$

用Z计算S个方程的总数，我们就得到

$$r_i q_{i,1} a_i^{m_i} \ln a_i = \frac{1}{s} \sum r_i q_i, a_1^{m_i} \ln a_i, \qquad (17)$$

但是方程(17)的右边不受i的支配，所以我们得出这样一个结论，最佳分配的条件是每个i和j具有相同的简单质量条件

$$r_i q_{i,1} a^{m_i} i \ln a_i = r_j q_{i,1} a^{m_i} \ln a_i, \qquad (18)$$

我们可以把方程(18)表示的结果用于整年的分配上。这种算法只是为了找出这样一些练习，以便通过这些练习的完成到学年末时能最大限度地提高学生的期望分数。我们已着手进行这种算法的推广和应用。其结果十分复杂，不便在此详述，而且在完全成功的应用之前还有些重要的概念问题尚须弄清楚。

为了提请大家注意其中一个紧迫的概念问题，我们将集中指向重要的方面。正如方程(18)所表明的，为了确定在什么地方进行下一个练习的时间的最佳分配，我们需要知道某一类的一个练习其一个错误的现行概率q_i——正如根据某一个目标或子目标所表示的那样。由于我们的分析是根据由子目标所表示的细致的分类进行的，所以我们就不要超出这些范围，忽略那些把子目标群集成束的目标。

现在的问题是如何估计q_i值。如果可能得到无限的测验数据，一个直接的解将是可行的，遗憾的是，在每个课程年度中有近200个子目标，因此每周都得到q_i的现行估计是不现实的。我们解决这个问题的方法是使用拟贝森法。首先，我们根据以往有关小学数学教学的大量经验指定一个先验概率分布和一个平均数。第二我们请尼加拉瓜的课程设计者和协调人给出她关于期望学生达到

的教前和教后成绩水平的估计值。这些估计值采取了正确反应比例简单概率估计的形式。我们还请她对每个子目标做出估计，然后我们再检验有测验数据的一定数量的子目标。

表4.7表明了在教学前所测得的10个二年级子目标的结果。正如从表中可清楚看到的那样，这些估计值对所有子目标来说都明显地系统偏低。在表4.8中，我们表明了16个二年级子目标的估计值与教学后所搜集的测验数据之间的比较。假若这样，正如表中最后一列所表示的那样，实际测验数据与估计值相比没有系统的倾向，但是在许多例子中有很大差异，实际上对一年以上的子目标来说这种差异大于0.10。

表4.7　　教学前10个子目标的估计值与测验数据的比较

子目标[a]	教学周	测验周	p的估计值	p的测验值		估计值与测验值之差
				全距	平均数	
VS4	7	6	0.30	0.46—0.80	0.66	0.36
OM1	7	6	0.10	0.24—.076	0.53	0.43
SP2	7	6	0.20	0.83—1.00	0.94	0.74
SN3	8	7	0.30	0.73—0.83	0.76	0.46
OS3	9	8	0.20	0.69—0.71	0.70	0.50
HS4	9	8	0.15	0.45—0.78	0.59	0.44
HM	9	7	0.10	0.35—0.95	0.70	0.60
HS3	10	7	0.25	0.53—0.63	0.59	0.34
OA8	10	9	0.35	0.53—0.75	0.63	0.28
VA9	10—11	9	0.10	0.24—0.46	0.35	0.25
		平均数	0.205		0.645	0.440
		标准差	0.093		0.154	0.148

a　这些缩写字及数字的含义请看302页的注释。

这些差异对于决定用来进行最佳分配的q^i值来说提出了难以

处理的概念问题和统计推论问题。值得注意的是，当问题是吸收课程设计者所给的非频率性的估计值时，如表4.7和4.8所示，关于贝森推论的文献也不能提供更多的帮助。贝森推论自始至终是根据频率数据进行的。采用如此蹩脚的工具来吸收概率估计值，以至于不依据一个新的后期形成的频率，这几乎是贝森强调主观频率的一个反论。

从表4.7所示的情况至少可清楚地看出，能估计出一个系统的修正值，并用来修正使用方程(18)所需要的qi的估计值。这已通过直线回归做到了。详细情况这里不多谈。这种情况对表4.8所示的数据没有太大的帮助。可以看出一点点倾向，pi的估计值高于pi的实测值，但这种倾向不太强，而且在测验数据和pi的估计值之间的关系上似乎没有可能进行可用的系统概括。

必须强调，在这部分所提出的整体分配模型的功效依赖于有个适当准确的qi(或pi， 因为总是$=qi1-pi$)估计值，并且还依赖于学习速率这些类似的估计值。进行课程改革的设计以及为完成收集和分析数据这种任务提供适当的人力、物力不是不可能的，但很难摆到适当的位置上。

表4.8　　　教学后16个子目标的估计值与测验值的比较

子目标a	教学周	测验周	p的估计值	p的测验值		估计值与测验值之差
				全距	平均数	
NC7	3	5	0.70	0.81—1.00	0.90	0.20
OA2	3	5	0.80	0.64—0.95	0.80	0.00
VA2	3-4	5	0.90	0.89—1.00	0.96	0.06
WN3	3-4	6	0.70	0.07—0.69	0.48	−0.22
NC 8	4	5	0.70	0.90—1.00	0.95	0.25
NC10	4	5	0.80	0.56—0.75	0.64	−0.16

HA3	4	5	0.65	0.88—1.00	0.91	0.26
VA4	5	6	0.85	0.80—0.80	0.84	-0.01
OC1	5	6	0.75	0.14—0.69	0.36	-0.39
HS2	5	7	0.75	0.48—0.55	0.51	-0.24
WN4	6	7	0.70	0.56—0.70	0.63	-0.07
VA3	6	7	0.90	0.88—0.95	0.93	0.03
VS3	6	7	0.85	0.70—0.75	0.73	-0.12
VS4	7	8	0.85	0.71—0.78	0.76	-0.09
OM1	7	8	0.75	0.46—0.87	0.72	-0.03
SP2	7	8	0.80	0.42—0.66	0.57	-0.23
		平均数	0.778		0.730	-0.048
		标准差	0.077		0.185	0.183

a这些缩写字及数字的含义请看302页的注释

因此，我们在这部分论述中所举的这个例子在下述意义上说是不完整的。一方面，概念框架是清楚的，应用目的也是明确的，而另一方面，因数据的收集和分析很困难，所以到目前为止仍妨碍成功地应用。据我们看来，关于课程的系统考虑，特别是课程的整体组织只要不涉及到有关实验数据的严肃的系统估计，一般都已进行了，我们举数学广播教学这一个例子的理由是为了说明这种非经验主义的态度勿需奏效，同时要有充分的坚持性和决心，把整体心理学模型用于课程的详细设计和组织是可能的。我们希望将来能报告我们的研究结果，并把上述分析贯彻到研究结果中去，我们相信所提出的统计推论问题可以通过建设性的方法加以解决。

杨　琦　译

五、学习结果的评价

罗伯特M.加涅 雅各布G.比尔德

I.引　　言

教育上有几种趋势增长了人们对考查学习结果评价的逻辑基础和技术的浓厚兴趣，在教学大纲中普遍采用细致规定的目标；越来越多地使用了各种形式的个别化教学；以及把可说明性概念应用于教育尝试中，这几种趋势都要求在充分合理的基础上来评价学生的学习。其中特别值得注意的是关于标准参照测验概念的讨论，这一概念是由格拉塞(1963)提出而由克劳斯(1963)加以发挥的。有关这方面的技术资料经由唐龙(Donlon,1974)；汉布尔顿(Hambleton,1974)；梅 西 克(Messick,1975)；米尔曼(Millman,1974)；波帕姆(Popham, 1975)加以综述，开始具有了有可能使评价的基本原理为实际工作者和研究人员掌握形式和内容。虽然关于标准参照测量的文章因使用了一些新奇的而且常常是任意的术语而令人苦恼，但是，这些文章的合理内核对于学习评价的系统技术的阐述不断产生影响。这方面文章的基本观点是首先考虑学习了什么，其次考虑找出学习的个别差异。

标准参照测量的定义很多，现提出几种可供选择的表述，它们是目标基础测量(引自米尔曼，1974)；标准参照测量、行为参照测量(唐龙，1974)；范围参照测量〔海夫利(Hively)、马克斯韦尔(M-

dxwell)、雷贝赫尔(Rabehl)、申森(Sension)、卢恩丁(Lundin),1973〕。虽然他们本身也没有比较清楚地表明他们所提出定义的模式，但是其中有些特点是很有价值的。经常引用的是格拉塞和尼特考(1971)提出的定义：

> 一个标准参照测验是精心编制来进行测量的测验，这些测量是可以根据特定作业的标准进行直接解释的。作业的标准一般通过给个体规定应完成的任务种类和范围来制定的。测量是针对从这个范围内抽出的有代表性的任务抽样进行的，而这些测量又直接以每个个体的被试范围作为参照的〔P653〕。

看来，这个定义是经过认真推敲得出的，它无论从广度上还是从精确性上来说都是足以包括"目标参照"、"范围参照"和"标准参照"这些具体表述的含义。它对阐明评价"学习了什么"意味着什么的主要贡献在于它指出了这些测量必须根据直接参照一个确定的任务范围做出直接解释(梅西克,1975)。

除了这个定义之外，还有测量关于学习了什么这一想法的几个来源，对目前的思考有重要影响。格拉塞和克劳斯(1963)提出关于熟练测量基本原理的发展，他们把熟练测量定义为："标准行为的评价——根据特定标准所给作业或输出的信号特点来决定〔P240〕。"这些作者特别关心测量形式的描述，这种测量主要是针对个体作业的当前状态而不是针对一种所推论出的品质或基本能力进行的，也就是说他们所设想的评价技术能测量标准作业本身的熟练，相比之下能力倾向测量是设计来预报标准作业的。

关于学习结果评价的第二个概念，由布鲁姆和他的同事(1956)、克拉斯沃赫尔(Krathwohl)、布鲁姆和梅西亚(Masia,1964),最近又有布鲁姆、哈斯丁(Hastings)和马道斯(Madaus,

1971)所描述的关于学习目标分类的概念。他们提出并举例说明了学习结果的类别，想象出所谓认知的、情感的和心理运动的一般学习结果类别的细微差别。这些重要著作提出的主要观点是必须根据不同种类的习得能力来设计和解释学习结果测量，而不只是根据代表不同成就程度的分数分布来设计和阐明学习结果的测量。

为详述"作业标准"而确定任务范围这一想法（正如格拉塞和尼特考定义所反映出的），已由海夫利(Hively)和其他一些作者进一步发展了(1973)。这些研究者知道去描述那些规定项目范围的极为确切的方法。任务范围就是具有确定的刺激和反应特点的项目形式，并且根据设计规则规定其构成项目的替换式。因此，这就为确定那些选择来评价已知学习目标的个别项目的代表性和适当性提供了合理的基础。尽管，有些研究者认为这些方法太麻烦，不切合实际（波帕姆，1975），但是，这些方法的重要性在于显示出规定任务范围意味着什么。

由于我们认识到了这些基本贡献的紧迫性，我们希望对当前关于学习了什么的评价或学习结果的评价的基本测量原理和技术作一综述。我们想到了学校学习中，训练计划中和其他所组织的教育事业中，可能产生的全部能力和素质。我们打算把以前所引证的许多有贡献的思想，以及其他一些想法合并到这个综述中，以便服务于这一中心目的。需要考虑的一些问题如下：

1.学习结果类型有哪些？

2.评价学习结果的质和量的测量其重要性是什么？

3.关于作业评价"直接性"的测量之间的差别是什么？

4.什么是"可直接解释"的测量？

5.怎样从任务范围中抽取项目？

6.评价测量的"适合性"具有什么意义？

7.表达测验结果的方式有哪些？

Ⅱ.指定任务范围

我们假设设计各种测量来评价学习结果的过程是从精心规定学习目标开始的。许多作者，另外还包括马杰(Mager,1975)，波帕姆，(1975),加涅和布里格斯(Briggs, 1974)都描述了明确定义学习目标的各种方法，一致认为在形式上适合测验项目派生的目标定义，至少包括以下几种成分：(1)一种刺激情境的描述；(2)一个表明预期能力的动作动词("主要"动词或"主要意图")；(3)动词所涉及到的一个或多个宾语；(4)一个描述动作是如何完成的动词或动词词组(如"通过书写")；(5)所呈现的工具和环境限制。

学习评价程序的设计者指定任务范围的第一步是从对学习目标给予适当定义开始的。这一步首先要回顾一下学习目标的定义，以便决定这种任务范围指定所反映的这个定义的准确度。这样就决定了任务范围的效度，同时展望所关心的学习结果评价项目的派生。为了达到这个目的，任务范围的指定必须尽可能准确地描述适合于所期望的测量的种类。

任务范围的指定比起学习目标来需要更详细的描述，海夫利等人(1973)所提出的使用项目一形式的方法已经提过了。波帕姆(1975)建议以极大目标的形式来进行任务范围的指定，这包括刺激成分、反应替代、改正标准和格式详述。在下面的部分，我们将尽可能全面地描述指定任务范围的基础，我们将要提出的一个重要建议是关于任务范围的决定最好早做出，而其他的决定可以晚些做出。我们试图根据以下几点考虑来指定任务范围的成分：(1)学习结果的类型；(2)质与量测量的目的；(3)测量的直接性；(4)作业标准。从我们的内心来说，至少想按上述几条的大致顺序来清楚地阐明任务范围的指定。

A.学习结果类型

指定评价项目的范围，首先要认识到习得了什么是分许多类型的。这些学习结果的种类不是由课程内容决定的(比如,自然科学、数学、历史)。最好把这些学习结果想象为人类记忆内容的各种知识，学习者获得并储存了许多不同类型的记忆组织。当我们设计各种抽样检验人类作业的测验时,正是试图为这些组织找出"指示物"或"探针"。因此，评价的设计一定要特别考虑到人们所要探测的记忆组织的类型。例如，如果所要测量的学习结果是投篮的技能，那么人们就应选择一种能适当表示这种任务的刺激与反应的测验,而不用那种需用语言回答有关球篮维数问题的测验。决定学习结果的类型反映了这样一个基本意图，目标所规定的正是将要习得的。

加涅(1972,1974,1976)对五种主要的学习结果进行描述:(1)智力技能;(2)语言信息;(3)认知策略;(4)运动技能;(5)态度。可把这五种学习结果想象成是代表不同种类的记忆组织的。这五种学习结果为了达到最佳能力状态，需要不同的学习条件。进一步说，正如所料，这些不同种类的记忆组织确实在人类的作业中显示出来，因此需要各种探查评价。

我们就这五种学习结果展开讨论。在关系到记忆组织的这个特性上，这五种学习结果似乎与关于认知多样化的现代学习理论的假设非常一致〔阿特金森、沙费伦 (Shiffrin, 1968; 鲍沃(Bower),1975;金茨奇(Kintsch)1974;鲁梅尔哈特(Rumelhart),林德赛(Lindsay),诺曼(Norman),1972〕。但是,我们看到其他分类也同样可以在各种学习结果之间做出必要的区分。这些分类显然与布鲁姆等人(1971)所提出的认知范围、情感范围和心理运动范围很相近。显而易见,共同的目的是想把学习评价的目的描述成是为了

获得某种习得能力(或素质)的指示物，并把这些习得能力想象为可以从学习者作业的系统观察中推论出来的记忆组织。

正如加涅(1972,1974)所定义的那样，这五种学习结果如下：

智力技能——学习者为了表示他的世界，在按规则操作符号时所生成的作业中表现出的概念和规则，并且特别由语言和数学规则体现出来。举例，修改这个句子："叶子将要柔和的摇动。"

言语信息——关于世界的信息，这种信息是以能用言语叙述的命题那样被检索出来的方式贮存的。举例：美国宪法第一修正案，禁止关于言论自由吗？

运动技能——在以平稳和计时精确为特点的运动作业中表现出的活动的顺序。举例：打印出字母R。

认知策略——学习者借以矫正自己的注意、学习记忆和思维诸内部过程的执行控制过程。举例：当回忆一个人的名字时，构画出一个能被很快恢复的富于幽默的视觉表象。

态度——获得性的内部状态，通常被想象成既是有认知成分，又具情感成分，调整着个体对各种客体、人物、事件的选择活动。举例：愿意欣赏古典音乐。

有关标准参照测验的著作中处处强调关于习得了什么的测量，这将导致人们去适当地考虑和阐述这五种学习结果。格拉塞和克劳斯(1963)的文章是这一领域的一个精华。这些作者举例所用的各种作业多到足以包括所有种类的学习结果。另外他们正对立足于象雷达操作、器材保养、解几何题、开药方、领导者的举止以及问题解决的智慧等各种任务测验要求的分析。

但是，在为学习评价广泛设想的基础与现有的关于标准参照测验文献中所看到的例子之间，有一个明显的对照。虽然有几个例子是关于语言惯用法的，但是实际上所有的例子(如，希维利等，1973)都是来自文学领域的。标准参照测验编制的原则几乎完全

是在智力技能方面得出的，很少注意到其他类型学习结果测量所要求的测验特点(波帕姆,1975)。

关于"标准参照"特殊意义的争论常常以选择来说明学习结果的具体例子为转移，例如,希维利等人(1973)就用"写出两个因数的积来完成一个乘法等式"这种例子来描述智力技能的评价。相反,有一些研究者〔唐龙,1974;埃贝尔(Ebel),1970;特劳伯(Traub),1972〕使用了言语信息评价的例子,比如象教师自编测验要求学生"回答关于美国宪法的十个问题"的百分等级这样一些言语信息,这两种结果反映不同类型的学习能力(记忆组织)。基本要求是决定怎么才能编制出最好的测验，以便评价出就这两种学习结果来说习得了什么。没有一种先验的理由能做出这样的假设，两种评价所要求的测量具有相同的形式特点，而不管它们是否都可称作"标准参照"。

学习结果和范围指定。为各种类型的学习结果指定任务范围，看起来是一个值得满足的要求。如果关于学习评价的想法和概念只能用来评价象数学和语法这样一些学科的作业，那么其价值就太有限了。虽然象这样一些智力技能无疑是学校课程的基本成分，但是，学校学习的广泛度和多样化也使得正确测量其它类型学习结果的方法变得同样重要。

除了智力技能之外，其他领域学习评价规定的叙述提出了许多挑战。在言语信息这种情况下，必须回答这样一个问题，学习的目标是获得"较广"的知识，还是"较深"的知识，或者是兼而有之。当学习者学习关于大萧条时期的历史教材时，希望他获得广泛的知识呢，还是有一个深刻的理解。在描述任务范围时，做出"高分意味着广而深的知识"这种假定，并不能解决这个问题。目前的情况就是如此。在其他方面的学习结果上也同样面临着定义和概念化问题。如何评价运动技能的准确性和平稳性的增长？辛格

(Singer,1975)描述了许多可供选择的方案。似乎有可能把这些评分办法与确定范围的维数联系起来。现在再来谈谈态度测量，这是一个很迷惑人的问题，也就是说在学习评价这个框架中，如何来想象这些习得的素质。能根据选择哪些项目来产生直接可解释的分数以描述态度范围吗？虽然特里乃迪斯(Triandis,1964)和格特曼(quttman,1969)的著作中给了一些有希望的提示，但是这种可能性没有得到明确的证实。第一步好象应当是描述态度的操作定义，反映出测量所用的方法。关于合理地测量出问题解决和创造性思维中所包含的认知策略，具有类似的困难。例如，如何能指定一个客观确定的范围来反映出作为一种学习结果的“首创性？”大体上说，如果关于习得了什么的评价被证实为是一种特殊的和有用的概念，那么似乎需要发展更新的技术。

B.质与量的维数

关于学习结果范围定义的一个基本概念，是作业的质与作业的量之间的区分。在传统的成就测验中，人类作业的这方面常常假定是等同的。因此，回答出关于波士顿惨案的十个问题中的八个常常被假定为象回答出十道整数乘法问题中的八道一样，同是一种测量。实际上，涉及到了任务范围的两种不同的维数。关于波士顿惨案的项目得80分，揭示了知道多少这个事实——也就是说，有多少个有关命题已存贮在学习者的记忆中。而在乘法问题中所用的这个相同的百分数是打算告诉我们整数乘法的智力技能已经学得多么好。为了把后者也理解为“多少”，就要假定学生分别记住了整数每个组合的积，这是一个完全不能接受的假定。

在学习结果的评价中，有时需要一个“量”的分数，也就是一个可直接理解为是“多少”的分数。在另一种情况下，一个值的分数(“多么好”)是表示探索什么。因为我们前边已提出了学习评价

必须要处理的五种主要学习结果，所以我们现在可以开始描述这些不同类型的学习结果之间对于作为任务范围特有的质与量的要求有些什么不同。

1.智力技能。当进行关于一种智力技能(规则或概念的应用)的评价时，所关心的是“多么好”的问题。某人希望测量两位数乘法或混合分数减法，或求联立方程根的能力掌握得多么好。用希维利等人（1973）所描述的方式可以完成任务范围的明确指定。这种方法包括了项目形式的定义、替代方案和以规定评价项目范围的广度和维数来表示的替换式。

注意到有证据表明单一的智力技能的学习是全或无，这具有一定的重要意义。这就是说，当一个单一智力技能的范围已被适当确定时，从那个范围中抽出的项目产生的分数要么都正确，要么处于偶然水平上——对“多么好”这个问题的回答是精通的或非精通的。例如，格雷厄姆(Graham,1974)发现，项目范围能以单一形式准确确定时，并于数学技能测验呈现出清楚的二项分布(如，计算利率和营业税)。但是，当所用的项目代表两维或多维的任务变式时(如，计算具有整数和分数解的行进时间)就发现分数分布于几个值上。霍威特兹(Herwitz,1975)得到了可与要求正、负数加法任务相比较的结果。这些发现再一次表明了对于能用单一项目形式范围(如－4＋9，－3＋7)规定的任务的全或无的学习和包含替换式(如加法－7＋3和3－7)变分的范围分数的差量。

所要求的智力技能的评价不是质的评价，而是量的评价。例如，如果一个学生已学过了分数，那么就可以去测验由这一主题构成的整群技能——分数加法、分数减法、分数乘法和约分等等。一个根据这个较大范围抽样的分数测验很容易编制——范围本身是由几种项目形式组成的，各自具有自己的替换式。关于这样一种测验的分数就能回答关于这个学生能完成“多少”分数的问题。

2.言语信息。有关言语信息(或"知识")学习的典型任务范围似乎是为了评价量的范围。例如，学生对于波士顿惨案，联帮论者论说或美国宪法条款知道多少？他对化学元素的特性或照明气灯的制作究竟知道多少？一般是通过设计那些从大概规定严密的范围内选出项目来回答这种定量问题的。这可以按照某一本教课书中某一段落或一段较长的原文材料的指定那样严密。或者可以不确切地规定为一个学生打算从他学习某一科知识所用的讲义、教课书或参考文献中学得的信息。但是，无论如何它是规定量的范围，并且只在最不确切(不正确)的意义上来考虑测量学生对某一科知道得"多么好"。

几个研究者〔安德森，1972；博尔马思(Bormuth)，1970〕提出的建议包含着这种可能性，知识的质能作为一种学习结果来加以评价。例如，有一个建议认为这种评价可以通过使用特定命题的语法和语义变化(释义)来实现，而那些特定命题是代表用来评价学习的知识项目。布鲁姆(1956)把认知范围分类为知识、理解、应用分析、综合和评价，也可以得出一个类似的解释。基本想法就是通过规定增加信息加工要求的范围来测量象知识的"深度"这种东西。这是一个引人注意的见解，其理论基础好象还没有完全搞清楚。哪种记忆组织能被看做是对应于言语信息的"较深理解"〔金特斯奇(Kintsch)，1974〕。为了有把握地设计出这方面学习的直接测量，在指定范围之前可能需要考虑进行其他的研究和理论发展。

3.运动技能。对于运动技能来说，最合适的任务范围是设计出来的可进行质的测量的范围。我们想知道学生完成一个特殊的运动动作有多么好——棒球他能击得多么好，或字母N他能打印得多么好。运动技能的范围通常能用合理的准确性来规定，而替换式能被用来表明所允许的工具和环境变化的限制。众所周知，大多数运动技能可根据准确性或速度来评价，并且一个项目形式要

求规定了这些因素。一旦范围的这些维数确定了，就似乎不难得到一个关于学习结果的质的测量。

当运动技能与某种运动或某种职业有关的情况下，可以指望进行运动技能的量的测量。于是，人们或许想知道打网球的技能(开球、反手击球、擦网险球等等)，获得了多少，或者修理汽车的工作掌握了多少。所描述的范围相应地包含复合测量的项目形式。正如格拉塞和克劳斯(1963)指出的，常常用一个核对清单来得到这种量的测量。

4. 认知策略。就所关心的学习评价的要求来说，对这种有趣的而又极为重要的学习似乎需要进行相当系统的思考。某些特定类型的认知策略，象背诵和记忆术的使用已经为相当简单的学习任务所认可(参看布朗，1975)；弗雷沃、韦尔曼，1976；曼德勒(Mandler)与史蒂文斯，1967)。但是，思维和问题解决所涉及到的许多策略还要进一步发现。显然，在为首创性问题解决这种能力规定范围时所需要的是关于"首创性程度"的测量——显然是一种质的测量。但是，一个主要的困难是搞不清楚这样一种能力是单一的还是几种不同的能力合成一个整体去共同影响人类的作业及其结果。例如奥尔顿(Olton)和他的同事奥尔顿、沃德罗普(Wardrop)、科文顿(Covington)、古德温(Goodwin)、克鲁特奇费尔德(Crutehfield)，克劳斯梅尔(Klausmeier)和朗达(Ronda，1967)在研究五年级儿童的创造性思维时就提出了这样的建议。

人们对识别、分离和测量若干不同的认知策略进行过各种尝试。有可能把吉尔福特(Guilford，1967)的大多数(如果不是全部)智力维数看做是只在规定特殊种类的认知策略。例如，一个象吉尔福特归类为图形系统认知的瑟斯顿旗子测验，可以被看做是评价一个关于旋转印刷图形视觉表象的策略。尽管吉尔福特提出了操作性的类别描述，但是选择若干测验项目所依据的范围不是可以

清楚解释的。在很多情况下，智力测验项目需要运用规则（智力技能）以及回忆信息，此外还要运用认知策略，这可能是他们无疑打算测量的。因此，他们进行的测量不是鉴别性的（加涅，1970），因为并不清楚它们所评价的是哪一种学习结果。这也同样适用于布鲁姆（1956）的提议，他认为评价问题解决能力的项目可分为“分析”、“综合”和“评价”。同样，在这些情况下，可能涉及到的策略没有与成功地完成项目所要求的智力技能分开。因此，不能说已经完成了关于认知策略范围的具有鉴别性的规定。

使用量的测量来评价认知策略似乎是不合适的，除非提出不能接受的假设，否则人们就不愿意把“解决了多少问题”的分数解释为关于策略数量的一个指标。但是，使用具有时间限制的多项测验，就有可能在所评价的策略效率的定义上进行质的评价。实际上，这可以是能力倾向测验所共同隐含的假设，也就是说可把这些测验看做是进行认知策略的测量。

5.态度。态度测量的问题很多，并且已成为几种评论性刊物的主题（吉埃斯勒Kiesler、科林斯、米勒，1969；麦克久雷Mcguire，1968；斯科特，1968）。一种态度可最明确地想象为是一种能影响行为习得的记忆状态。根据这种看法，可把态度衡量为在对某种客体、人和事件上选择个体动作的一致性（加涅，1976）。确定这些选择的范围可沿着几个维数进行认真的阐述〔特雷纳迪斯（Triandis），1964〕。在评价个体动作的选择上，例如，关于“与黑人的社会交往”特雷纳迪斯是根据以下两点所构成的范围来选择项目的：（1）社会交往的形式；（2）黑人的社会个人特征（职业、年龄等）。这种方法或变式都假定可用来确定范围，使直接解释态度的分数成为可能。

某一种态度（如对个人的废弃物处置的态度或对阅读现代小说的态度）的测量需要考虑这种态度的“强度”。象做礼拜的次数所

表明的那样，有些人对做礼拜态度强烈，而有些人态度消极。态度的强度可被看做是在正、负两个方向上变化着。这些考虑意味着应把态度的强度看做是一种质的评价，虽然质本身是从象动作选择的频率和一致性这样一些测量中推论出来的。

大多数现有的态度量表和工具是量的测量，因为它们依赖于从许多虽然逻辑上有关但却不同的项目范围(参看肖和赖特,1967所描述的各种态度测量工具)中所得的累积分数。因此，一般不使用实际是属于"对教师的态度"，"对同学的态度"，"对学习的态度"，"对放假的态度"和其他许多特定项目范围的语句构成的工具来评价对学校的态度。然后由累积分数来做为"对学校态度"的总指标。显然，根据这种工具所获得的测量是量的测量，其意义在于"以积极的形式回答了客体种类的多少个方面?"因为所获得的分数通常不能单独与确定的动作选择范围相联系，充其量不过是学习结果的粗糙测量。

6.任务范围指定的含义。识别出某种学习结果，并对这种结果进行适当的质的或量的测量，就能使评价程序的设计者确定出任务范围内容的一般性质。某种学习结果的选择通常是根据学习目标的分类，这一点是显而易见的(加涅和布里格斯,1974)，或者，换句话说，是根据其主要意图(马杰,1975;波帕姆,1975)。然后必须考虑并决定这种学习结果所提供的质与量的测量的可能性。这种决定影响着该范围项目分数的可解释性。例如，"智力技能"这一学习结果就要求对那种作为习得性实体的技能进行质的测量。这是不能从量的测量中推论出来的。相比之下，关于"言语信息"这一学习结果的测量通常是一种量的测量，但关于作业的质，一般不能很有把握地从这种测量中推论出来。

当然，在进行质或量的测量的任务范围之间做出选择还有其他理由，下边将要进行讨论。这些属于有意地使用了根据学习评

价所得的分数。因此，使用一个测验来综合评价某一课程的学习需要量的测量，把几个有关的任务范围合起来，其中每个范围本身相应地产生一个质的测量。

在表5.1中给出了为确定任务范围内容的设计决策总结。如表所示，根据追求的是质的测量还是量的测量来看，指定任务范围内容的性质可能是不同的。这种测量的适当性反过来又受到与所评价的预期能力和素质相一致的学习结果种类的影响。

表5.1　各种学习结果的任务范围内容依赖于质或量的测量选择

学习结果类型	测量类型	任务范围内容
智力技能	质的	明确规定项目形式和替换式
	量的	分别识别项目形式和几种不同技能的替换式
言语信息	质的	项目形式定义的不适当的推理基础
	量的	根据所回想起的一般的或特殊的命题("大意","具体细节"等）确定项目形式和替换式
运动技能	质的	根据作业的维数，包括准确性和速度来明确规定项目形式
	量的	分别确定与具体人类活动有关的若干套项目形式
认知策略	质的	确定象"效率""首创性"等作业测量的项目形式
	量的	不合适
态　　度	质的	对各种客体、人，或事件的个人动作选择的项目形式和替换式，根据种类属性明确确定。
	量的	有关态度的汇集，分别确定为一个质的测量；为每种有关态度分别确定的项目形式的集合

C、测量的直接性

任务范围指定需要的另一方面常常被认为是测量的直接性。

这种指定包括决定所定任务的适当的刺激情境和反应措施(希维利等人,1973;波帕姆,1975)。任务范围指定的这些成分可以从下例中所表明的教学目标的相应部分推导出来:

1.给出一个印好的3□6=18的等式,并说明需要补上恰当的运算符号(刺激情景);运用算术运算规则,填上适当的符号,完成这一等式(反应)。

2.给出印好的句子,其中把形容词用作副词,并印出需要改正这些句子的说明(刺激情景);运用副词的构成规则和习惯用法,在误用形容词的地方写出副词(反应)。

3.给一块"1×6"的木板,在板面上从一边到另一边画一条线(刺激情景);然后用手锯锯开木板,允许锯缝与所画直线有半度之差(反应)。

1.测量直接性的变式。有许多理由可能使教学目标(因此也是指定任务范围)的刺激和反应特点与"实际的"或"最终的"教学目标必须脱离。有关教育大纲中规定的近期目标和最终目标之间的区别,已由格拉塞和尼特考(1971)讨论过,而林德奎斯特(Lindquist,1951)更早讨论过。正如这几位作者指出的那样,某一学习课程的近期目标是在教学告一段落时能合理而可行地进行测量的目标。这些近期目标通过某种理性分析的过程,或者更难得的是根据经验证明与最终目标联系起来。因此,某一职业教育课程的最终教育目标可能是培养电视维修的能工巧匠,而近期目标仅仅是明确规定电视维修的各种胜任能力,这些胜任能力是可通过使用教学情况中现成的器材设备进行评价的。

甚至近期目标和由此推导出来的任务范围可能受到经费、时间限制,或危及受试者生命安全的这些可行性因素的约束。任务范围的指定可能因为这些考虑而需要测量直接性的各种变式。例如,学习基础力学的学生可能已经知道在提起各种滑轮组中的重

物时，需要找出必须使用的力的数量。在这种情况下，近期目标可想而知包含几种不同的“情境”描述。可能这样叙述道：“已知一个滑轮组……”，或“已知一个表明滑轮组的画图……”，或“已知一个滑轮组的打印的说明……”。显然，这些不同的情境如果反映在任务范围的描述上，将意味着在评价中具有不同程度的方便性和可行性。

林德奎斯特(1951)把测量直接性的程度描述为以下四类：(1)相同要素；(2)有关行为；(3)言语表达行为；(4)学生的知识。格拉塞和克劳斯(1963)识别出三种直接性：(1)专心工作；(2)模拟；(3)相关行为。这些类别与林德奎斯特的分类相类似，虽然增加了模拟作业测量的差别是比较重要的，并且由菲茨帕特里克(Fitzpatrick)和莫里森(Morrison，1971)做了进一步阐明。他们俩提供了许多表现多种程度直接性的模拟测量的例子，并且叙述了他们的评价特点。

2.关于测量直接性的效度。正如本文引用的各种类别所表示的那样，显然测量直接性程度的选择，在设计实际的测量中是相当有价值的。虽然从概念上说，测量直接性的问题是与所编制的测验的内容效度有关系的。好象是任务范围本身内容效度的合理证实，似乎最终必须根据如目标所示的任务定义的成分做出。评价设计者一般认为，任务成分不仅包括刺激情景和反应，而且还包括过程(布鲁姆，1971；克罗贝纳奇，1971；波帕姆，1975)。大概，任务范围的内容效度是通过详述在这所有三个因素上脱离开对所要求的目标在最小可能程度的描述上来达到的。

正象前几节所暗示的，我们对任务范围指定的处理，是把“过程”看作是由“主动词”(加涅和布里格，1974)或目标的“主要意图”(马格，1975)表明。正如一些作者指出的那样(例如克罗纳贝奇，1971；梅塞克，1975)，有可能过程的确定是一个结构效度的问题，

这种结构是从理论上推导出来的。这种观点至少与任务范围描述了不同形式的记忆结构的想法是一致的，正如目前十分活跃的各种现代认知学习理论所暗示的那样(鲍尔,1975)。

由目标所意指的任务范围与过程的脱离可能导致较低的效度。例如，按允许使用与填空项目相反的多方选择项目的方式指定范围，常常看作是涉及到一个过程变化。在许多情况下(不是全部)对多方选择项目的反应需要再认过程，而不是回忆所涉及到的全套过程。虽然再认和回忆需要相同的过程这个假设不能说不被证明是不成立的。但是，有确凿证据表明涉及到不同的过程，这一点是值得认真注意的(安德森和鲍尔,1974;克拉茨基,1975)。

在学习目标中，描述的刺激情境的表示的直接性对任务范围内容效度的问题有特殊的说服力。刺激是引起评价所要求的无论什么记忆内容的恢复和回忆的线索。因此，任务刺激方面的变化对于在评价情境中学习者的作业，显然有直接的影响。作业线索的变化可能不仅包括那些任务本身所固有的变化，正如当测量流体流量俩这个任务的言语描述时，可用一个实际的流体刻度容器代替那样。而且环绕着"测验情境"中的线索，象场所、到场的人物，外来的噪音，学习者的心理定势等等，通常都与"实际情境"中的线索十分不同(加涅,1975)。进行某种程度的模拟，特别是关于刺激情景的模拟，在评价中是不可避免的。

评价所要求的反应也可能不同于实际情境中的反应。但是，人们对反应变化本身的了解没有象对评价程序的效度有特殊的影响了解得那么清楚。举个简单的例子，写出一个句子回答言语信息问题与说出一个句子来回答同一个问题，好象不是进行不同的测量。根据某些作者的意见(参看奇斯勒等,1969)，态度分数与行为的直接测量之间的不一致，最好被看作是由引起个体反应的刺激情境的相当不同造成的，而不是由反应本身的差异造成的。就运动

技能来说，即使引起反应的刺激是突出的，也总是在反应速度和精确性的广泛变化中表现出高度的学习迁移(辛格，1975)。对指定反应保持高度逼真的关键，常常被看做是一个准确地识别活动所包含的过程，这样看是比较合适的。

D.作业标准

学习结果评价方面的许多著作都有把分数作为直接参照“标准”的想法。格拉塞和尼特考使用了“作业标准”这个词。波帕姆(1975)阐明了范围的描述应该包括“这样一个方面，即用清楚的语言指出是什么构成可以接受的答案”。海夫利等(1973)使用的项目形式提供了“记分的说明”。马杰(1962)在他的一本很有影响的书中把“作业标准”的成分并入一个教学目标的定义中。然而正象这个词本身所表明的那样，可把标准参照测验想象成为一种测量技术，它产生的分数可以直接作为一种作业的“标准”。

“标准”这个词有很多意思。正象格拉塞和尼特考(1971)指出的那样，在教育和心理测验的文献中，标准一般有两种不同的意思。一种意思说它是一个与某一测验分数相关的测量连续体，如词汇测验的分数就与在校成功的“标准”相关。另一种意思说它是一个作为成绩指标的可接受的分数水平，例如指明80%的项目是正确的。唐龙(1974)探讨了这两种意思之间的差别，他认为忽视这一差别是许多混乱产生的根源。

就指定项目范围来说，正象格拉塞和尼特考(1971)认为的那样，“标准”的这两种意思都必须予以否定。此处的标准是指定任务的范围，以便从这个范围中选出项目来进行评价。可以看到，这个范围指明一类任务，但这些任务本身并不安排在连续体中；因此也就不必要求这些任务产生的一个分数的连续体与某个其它测量进行相关比较。这个范围也不必指定一个“通过分数”，即完成

给定数量中的多少项目，被认为是可接受的。范围是一个"标准"，其意义在于它可以识别出一类项目，这些项目能清楚而明确地评价出习得了什么，如同学习目标所规定的那样。

如果把分数"直接解释"为学习结果，只意味着这些分数与某一不易指定的范围中的某些项目的成绩水平有关，那么这种想法就无多大价值了。正如一些作者〔唐龙，1974；埃贝尔(Ebel)1970；特劳布(Traub)，1972〕指出的那样，象80%或每分钟70个词这样一个掌握标准，它本身并不能使标准参照测验的分数比象IQ量表这种常模参照测验的分数做出任何更直接的解释。

一个正确性的标准，可具有关于任务范围的准确意义。因此可用它来做为每个项目作业的评分参照标准，而不是通过多少测验项目数量的评分参照标准。例如，穿针这项运动技能就要包括指定0.5mm的针眼作为一个正确性的标准。马杰(1962)的一个例子是把雷达距离指示器调整到八分之一英寸内的标准模板形式。所以正确性的标准涉及到任务范围所确定的一类任务所固有的特点，而并不涉及"一个掌握水平"的任意赋值，例如对0.5mm针眼穿针的掌握水平认为10次中有8次穿过就算成功了。这种标准在教学过程管理中可能是很有用的。但是它不是指定测验项目范围所要求的固有的标准。"正确性标准"为一个任务范围所描述的每个项目提供着这样一个标准。

Ⅲ. 测验的编制

编制测验来直接评价具体学习结果的过程在某些方面与设计各种方法来测量人的广泛特性或品质的经典过程是不同的。这两类测量之间的最根本的差别通常可归结为，可参照一个指定的常模组来解释个体作业。因为在常模参照测验中存在着人员方差这

个固有条件，所以形成的统计程序就假设这种方差是存在的。对于常模参照测验来说，采用了能使这些统计指标尽可能完善的测验编制程序。

然而，有许多学习结果用经典的常模参照模式并不能很好地测量出来。近二十年来人们显然越来越关心更直接地评价学习成果，而不需要参照组来解释的选择模式。由于缺乏编制和评价标准参照测验的理论指导，以致使测验设计者经常遭到挫折。我们不想在本书中对这个问题领域提出综合的处理。但是，有三个关于测验编制的问题似乎是十分重要的，我们将在这部分对此进行讨论。这三个问题是：选择测验项目，质与量的评价，测验的长度。

A.选择测验项目

为了选择一个标准参照测验的项目应考虑两个基本标准。第一，这个测验所包括的任务样本必须对所关心的更大范围内的任务具有代表性。第二，这个测验所包括的项目数量必须大到足以对个体作业进行可靠的推理。其中，第一个标准与测验的内容效度有关，第二个标准与结果分数的信度有关。

测验分数可根据大范围的任务所要求的作业来解释。测验任务的一个概率样本会很容易得到这样一个分数。奥苏伯尔(1968)指出，在这种测验模式中概括化基于从范围内抽取项目程序的操作定义。所以，测验设计者和测验使用者应关心范围的描述和抽取项目的程序，而不是任何一个测验的特点。

关于从范围中抽取项目的适当模式，大家达成的一般协议是，随机取样或分层随机取样。奥苏伯尔（1968)、米尔曼(1974)、海夫利等(1973)和克里沃尔(Kriewall)(1972)等指出在保证测验分数具有意义这一点上，随机项目取样的模式是符合需要的。当项目取样真正作到随机的时候，那么分数的绝对值可作为一个比率

测量来解释。也就是说，由于忽略了误差，就可以认为一个80%的分数表明它在这个范围中所能完成的任务数是一个40%的分数所完成的任务数的两倍。克里沃尔(1972)指出，只有项目取样确实是随机的时候，才能假设具有这种比率的性质。而把一个有目的的项目抽样看成是随机的，这种普通作法会得出关于未知效度的分数。此外，项目选择有很大的判别指数和难度值的偏分布可导致类似的效度问题。

当一个范围已被明确地规定下来，并且从这个范围中选择项目的程序也被设计好的时候，就能很容易地形成平行的测验形式。这种测验编制的模式与利用相继平行的测验形式来达到预测和复测的目的的各种教学系统的需要是高度一致的(克里沃尔，1972)。

B.质与量的评价

某些关于标准参照测验的著作讨论过某种学习结果所特有的问题和程序。但没有提出合理地进行概括化的清楚界限。对所有标准参照测验的具体问题处理得过于概括化可能是引起混乱的根源之一，因为过于概括化会忽略测验问题的不同类型及其解题策略的差异。我们认为评价学习结果的方法在很大程度上是受学习结果的类型和有关任务的质与量的特点所制约的。

正象在第Ⅱ部分讨论的那样，量的测量要回答"多少"的问题，测验的结果应该对正确完成这个范围的任务的比例作出一个估量。例如，一个拼写测验的结果可以表明能够正确拼写出某一个表中词的比例或某一类型的比例。我们希望这种测验的成绩在许多分数点上是不同的。标准参照数量测验项目的选择和其结果的解释所使用的策略显然与常模参照测验所使用的策略不同，虽然，大部分标准参照测验量的分数的统计与常模参照测验所用的统计相类似，但项目的指标和测验适合性是不同的。(但是，应该记住

常模参照模式可能适用于某些量的评价。)

相比之下，质的测量能告诉我们，一个任务完成得“多么好”，对于质的测量来说，把分数分为掌握和非掌握两类最为有意义。我们发现单维智力技能的质的测量显然会产生双峰分布。这些研究的数据表明分数分布的状态依赖于目标的复杂性(或维数)。例如，格雷厄姆(Graham, 1974)编制了一个标准参照测验，测量下述目标的熟练：

“给出两点间的距离和运行速度，让学生决定所需要的运行时间。”

根据下面的格式产生各项：一辆汽车行驶了d公里，平均速度为每小时r公里，多少小时可以走完这段路程？

答案：________________

我们备置了关于这个目标的一个十项测验(A型)，其中有5项是整数解，另外5项是1/2小时的分数解(例如，$1\frac{1}{2}$，$3\frac{1}{2}$小时等等)。对558名中学生进行了这项测验。结果分数的一个次数多边形，如图5.1中虚线所示。

这个分布显然是三峰的。测验明显地把学生分为三个组：(1)这组学生能够完成整数解和分数解(众数为10分)；(2)这组学生只能完成整数解(中间众数为5分)；(3)这组学生哪一种解也不能完成(左边众数为零分)。为了验证这个假设，格雷厄姆又设计了第二种测验(B型)，解全部为分数，仍由上述学生来完成。“全部为分数解”的测验次数多边形，如图5.1中的实线所示。中间众数为5分的分数在这个分布中消失了，只表现出一个关于分数的双峰分布。低端众数的次数增加了，大概包括了许多在A型分布中处于中间众数的学生。第二套测验的项目更接近于单维测验，分数的分布是双峰的。数据表明大部分学生或者掌握了技能，或者是没有掌握

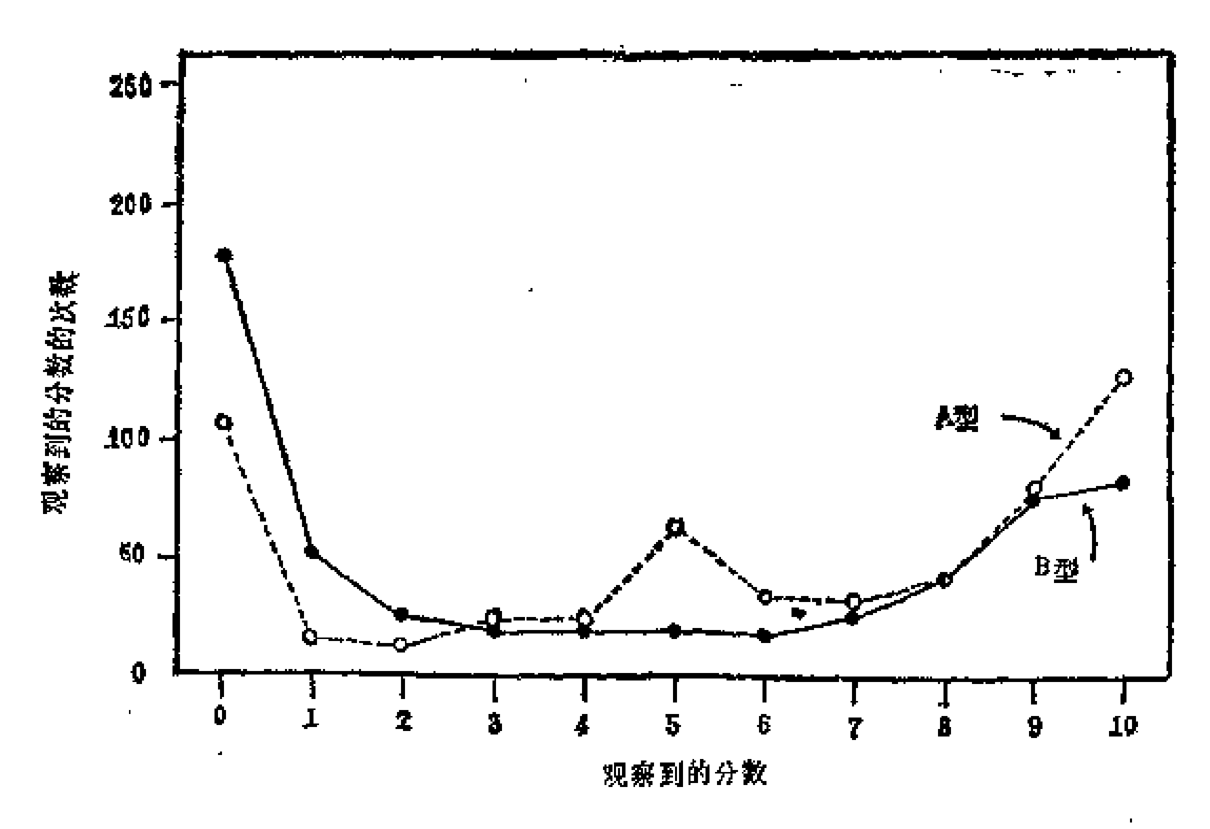


图5.1　计算运行时间的标准参照测验分数的分布
(A型题有整数和分数两种解,B型题只有一种分数解)

技能，几乎看不出部分掌握。这种分布的分数要比矩形分布或正态分布的分数更适于两分法表示。这些数据及格雷厄姆和霍威斯提供的其它数据都清楚地表明目标的维数决定着分数分布的形状，因而也就决定了两分法分数的可解释性。也可以从理论上看出，任何一组非常同质的项目，其近似的单位元素交互相关，因而可以产生一个关于总体分数的双峰或两点分布。

“量”和“质”这两个术语似乎有效地表达出这两种测验类型的主要差别。可把质的测验设计成能用准确定义的技能知识来直接解释的测验。这些测验的项目应该是同质性的或单维的。然而，也有证据表明对项目维数的主观判断往往是不可靠的，而通过经验来证明的这些项目的同质性好象更符合需要。非常同质性的测验可能相当短，但能提供可靠的结果。

量的测验也打算设计成能根据一个规定的任务范围的熟练性来直接解释的测验。量的测量项目不能象质的测量那样需要具有高度的同质性。这些测验分数的分布形状取决于范围的维数，也就是取决于选来表示范围的那些测验项目的同质性。

C.测验长度

标准参照测验长度的问题已被证明是一个最难回答的问题。有许多测验方案针对每一个目标仅用了一个项目，而有些研究性的方案则用了多达30个项目。虽然，在开发能回答这些问题的标准参照测量技术方面做了一些工作，但所做的努力微乎其微，以至于波帕姆(1974)称这个领域为"技术荒原"。

常模参照测验的长度一般要兼顾考虑实施测验所需要的时间和测验内部一致性的信度这两个方面。由于信度是项目数量的一个函数，因此所编制成的测验要长到足以能提供可靠的分数，但也不能长到需要多课时来完成。折衷的结果是所编的测验通常具有20—50个项目。从这套项目中获得的信度系数的平方根可解释为获得分数与真实分数之间的相关。

当进行标准参照测验时，获得的分数其方差常常很小或没有方差。在这些情况下，传统使用的信度和效度指标可能是不确定的或无意义的。因此，就提出了一些评价标准参照测验适合性的替代程序，但是到目前为止，还没有一个程序可以做为一个测验长度的适当的决定而被充分肯定。阐明标准参照测验长度的一般性原则可能会受到现有的各种标准参照测验种类和测验应用的限制。在这个描述性的标题下，你可以发现从一个项目的诊断性测验到一个完整课程的最终考试的各种测验。测验应用的多样性也使得用一个方针决定测验长度是很困难的。

一种得到广泛研究的应用是那种能通过因材施教(IPI)促进

学生进步的测验。米尔曼(Millman,1973)研究了这种测验的长度问题。他把在测验中正确的百分数分数看成是在这个项目范围内学生正确完成测验的百分数的估计值。然后利用二项式概率表评价那些该失败而错误地通过了和该通过而失败了的学生在每一个真实作业水平上的百分数。米尔曼表对这种测验的误差性质，包括测验长度之间的各种关系，分割分数，和由每一种错误分类引起的误差百分数都提出了一定的见解。但是，他的工作似乎没有提供出测验开发者们一直寻求的决定测验长度的简单决策性的程序。

诺威克和刘易斯(Novick、Lewis,1974)指出米尔曼表提供了关于一个学生用其真实功能水平获得某一测验分数的概率。但这种方法的应用受到了限制，因为每个学生真实的功能水平是不清楚的。他们提出了一个说明测验长度的框架，它始于学生的测验分数，这些分数是已知的，并提供了学生真实功能水平超过一个规定的标准水平的概率估计值。用贝尔斯(Bayes)定理把对学生预期的作业知识(对具有类似训练的学生作业事先了解)与学生测验分数结合起来，以便提高正确分类的概率。他们也论述了由正失灵和负失灵误差产生的相对损失的差异问题。例如，他们认为，使一个失败的学生得到提高所付代价是使一个通过测验的学生保持成绩所付代价的两倍。处理不同的损失比率程序在他们的论文中做了描述。诺威克和刘易斯为选择掌握水平的标准，学生功能水平的先验分布和有关假设提高，假设保持的相对损失提出了一些具体的测验长度的建议。这些建议是认真拟定出来的，对任何一个面临着测验长度问题的人来说都具有现实的意义。

我们认为在学习的主要类别之间，特别是在言语信息和智力技能之间，以及质与量的维数之间加以区分，会使决定测验长度的实践问题变得稍简单一些。例如，单维智力技能的质的测量可

以非常短，甚至短到只有一项。量的测量的长度因依赖于任务范围的同质性可能要相当长。无论是质的测量还是量的测量进行适当的评价所需要的长度都得考虑到米尔曼(1974)、诺威克和刘易斯(1974)的论述，并依赖于测量的应用目的。另外，长度的决定还应考虑到猜测是否是测验中的一个因素，并且根据分数做出个体还是团体的决定。

我们主要讨论了根据学习者的作业来评价学习成果的问题。在通过一个个别教学程序或者一个掌握学习的程序来控制进步的时候，使用这种测量通常是以按层次安排内容为前提的，因此，熟练掌握一个教学单元的目标将为促进学生进入下一个任务的学习提供非常有用的信息。关于内容结构的各种替代假设可以导致人们去探讨其它观点的控制作用。这些可供选择的观点实际上可能影响着测验的设计。

Ⅳ. 评价测量的适合性

测验分数一般以一个大范围的任务样本为基础，并提供关于这个范围熟练掌握的概率估价。这些估计值与该范围真实值的接近程度是测验使用者最关心的。测量理论的一个主要部分是处理获得值和真实值之间的这种差别，即测量误差。

通过理论调查和实验的研究确立了适合于不同的量表编制模式的标准。这些研究得出了测验结果可以接受的标准。然后再把这个标准按常规用于测验应用的结果中来确定和证实分数的适合性。直到出现了标准参照测验，才有了许多具有适合性的评价测验，包括大家公认的各种效度和信度的程序。效度可细分为内容效度，结构效度，并存效度和预报效度这些较细的程序亚类。除了内容效度外，相关系数是效度的共同指标。测验信度是由重复进

行同一测验或替代形式测验所得分数之间的相关或由一个内部一致性的系数表示的。只有当测验的效度和信度达到可接受的标准时，才能认为测验是适合的。

以积矩相关为基础的效度和信度的传统指标，在测量的方差接近零时，就无法确定了。但是波帕姆和赫斯克(1969)指出，当标准参照测验用于教学环境中时，无方差的分数分布实际上正是所希望达到的。例如，我们期望所有的学生都能在一个设计好的程序教学单元之后，成功地完成后测。如果出现这种情况，那么所有的分数都相同，方差就是零，但内部一致性的α或KR—20指标就无法确定，而且与其它变量的相关等于零。当这种分数分布出现时 测验的适合性的传统指标是无效的。

现在已经指出了许多标准参照测验适合性的指标，布伦南(Brennan 1974)和米尔曼已对这些指标做了综述。现已经开发了许多技术来评价制定某些教学决策所用的项目和测验的适合性，例如，通过一个教学程序来控制进步，但这些都不是我们在本章所涉及的主要问题。而我们所注意的是，在比较一般的和不受限制的意义上，学习成果测量适合性的适当指标。

A.效　度

不论对质的还是对量的测量来说，最适合的效度类型是内容效度。内容效度是通过测验范围的指定过程和适当的项目取样来确定的。首要的目标是测验项目对任务范围来说要具有代表性。各种项目形式(或阐述的目标)的使用和随机取样程序为内容效度奠定了基础。只有通过清楚地叙述范围指定和取样程序才能明确这种效度。

可能有必要象考虑内容效度那样去考虑结构效度。我们已经提到过，范围指定和测验编制应该以目标和有关质或量所规定的

学习结果的类型为基础。寻找实际证据来支持关于在完成各种测验任务时涉及到不同的记忆组织这一争论，就构成了结构效度。

B.统计模式

这里所提到的评价学习结果的理论模式实际上是一个项目取样的模式。洛德(Lord)和诺威克(1968)讨论过这个模式的长度，克里沃尔(1972)重申了标准参照测验的来龙去脉。构成这个模式的基本思想是相当简单的。这里有一个关于项目的假设范围。某一个个体能够对这个范围内的一部分项目做出正确的反应，对这个范围的项目做出正确反应的百分数就是学生真实的分数。测验是由这个范围内的项目的样本组成，因此测验结果的正确的百分数乃是整个范围内正确反应百分数的一个估价值。

这个模式主要的统计特性是：(1)所观察到的测验分数是估价真实分数的一个公正的估计量和一个充分的统计；(2)对一个个体来说其正确数的误差方差是：

$$\sigma^2_{\text{量数}}(a)=\frac{1}{n-1}x_a(n-x_a)$$

在这里

x_a=个体a的获得分数

n=测验中的项目数*

除了在分母中使用n－1校正样本的偏倚外，这是一个双峰分布(npq)方差的习惯表达式，其中n是样本观察数，p是某一事件的概率，q是1－p，所以，测量误差是由测验中的项目数和被试的熟练水平所决定的。当测验项目较多，熟练性极高或极低时，测量误差就

* 原文中为text，疑误，应为test

较小。

可以说熟练性的适当测量是根据一个从零到1的量表所测得的一个正确反应的概率。这种熟练性测量(对一个人)的标准误表示为

$$\sigma_{\text{量数}}(a)=\sqrt{\frac{pq}{n}}$$

在这里

p＝已知范围项目的比例，通常根据测验正确的比例来估价

这个系数是一个与测量有关的误差量相适合的指标。信度指标，例如，α，测验一复测或平行形式都直接与测验被试组的方差有关。这些指标为复测组的相对位置提供了预期稳定性，但是这些指标对不需要参照其他人的标准参照测验来说是无意义的。我们认为上面描述的测量的标准误是充分的。

V．测验设计所需要的决策

在本章所述的概念框架中存在着一系列在测验的选择和设计过程中要回答的关键性问题。在设计的早期阶段认真考虑这些问题可以提高所编制出的测验的适合性。表5.2给出这些应认真考虑的问题和一些可供选择的答案。表中的问题是按着通常的处理顺序排列的。这样排列只是为了方便，不是固定的。在某些情况下，可能需要同时考虑两个或更多的问题。而对某些问题的回答可能限制了可为后一个问题使用的选择。由于把这些概念应用到不同的评价问题中，所产生的测验描述可能会有很大差别。

我们认为设计一个测验最初的决策应该是把学习目标划归为前面描述的五种学习结果的一种。第二决策就是要决定是进行质

的测量还是量的测量。例如，是对某一个单维智力技能(如使用前缀anti—)进行质的测量呢，还是进行关于"前缀和后缀的使用"的量的测验?这样，测量的直接性也就成为一个决策性的问题了。你希望以直接的方式测量呢，还是使用一个任务的模拟更好?这些决定自然而然地就导致下一步的测验设计(表5.2中的D)，根据项目范围的大小和替换式确定项目的范围。

指定了项目的范围，才有可能进一步考虑项目取样问题。要针对这个范围内的每一个目标和子目标进行取样，从有明确的替换式的项目形式中取样，取样时要特别注意内容效度。如果所设计的是一个量的测量，就要特别注意在采用分层随机取样时其内容的代表性。项目的数量问题首先要考虑到由什么构成了测量的可接受的标准误差，接下来要考虑分数的使用问题。一旦得到了测验结果，你就可以着手选择最符合预期使用的分数表达方式。例如，为达到诊断目的使用的分数，一般要有能表示具体范围或子范围作业质量的百分数分数的形式。而对一门特殊课程的量的测量所获得的分数既可以用百分数表示，也可用百分位数或与常模有关的其它测量表达。

A.测验设计的例题

举两个例子来说明如何根据表5.2所给的一组问题进行一个测验设计。这些例子通过依次考虑这些问题并对测验设计提出可能的回答，来说明一下测验概念化的过程。

例1. 第一个目标是从佛罗里达州8年级测验程序的"基本数学技能"部分选出来的。

"标出一个商品的价格和购买商品使用的货币单位，要学生计算找回来的零钱的总数。"

表5.2所示的一系列的问题和可能的答案如下：

(a) 学习结果的类型是什么？

学习结果是智力技能，而且能够同其它四种学习类型清楚地区分开来。并且需要一些言语信息来完成任务(货币单位的名称)，但我们假定这是已知的。值得注意的行为是对不同的价格和货币单位值反复地运用计算零钱的规则。

(b) 是量的评价还是质的评价？

从测验结果中得出的基本推理关系着作业的质，而不是作业的量。也就是说，目标在于要找出被试计算零钱完成得多么好。而很少关心一个被试能正确地完成一组计算零钱问题的40%，另一个被试能完成60%，因为这两个人的推理，可能都不是完成这个任务所需要的恰当的推理。

(c) 如何进行直接测验？

学生在实际购买商品时，最终作业是很有意思的，然而"言语描述"这种测验格式的效率也是吸引人的。最终必须根据对"言语描述"反应和对"实际"付钱情景反应之间的关系来决定问题。在这种情况下，选用"言语描述"是因为我们假设当通过言语描述来学习时，这种技能会很容易地迁移到实际情景中去，而且由于刺激性质的变化而产生的错误不是特别重要的。如果所评价的情景是教学情景的一个组成部分，那么，就可以选择"模拟"。

(d) 如何描述范围？

"项目形式"的处理应该符合目标。这个测验任务的范围应该由一个具有变量因素替换式的项目形式来确定。下面例题是海利夫(1973)所描述的程序的一种比较简单的变式：

题目形式：

如果你买东西所花的钱是_x_，而你付给店员_Y_美元，问应找你多少元钱？

替换组：

x＝＄0.1到＄99.99中的任何一个量。

Y＝1,5,10,20,50,或100(Y>x)

(e) 如何进行项目取样?

一个测验项目是随机地从替换式x和y中选出的一组值组成的，用这个值代替项目中的适当空格。重复这个过程可以组成许多项目。所组成的这组项目又构成了这个任务或项目范围的一个项目随机样本。这些程序就保证了如此构成的所有项目对目标的有效测量。所选用的项目数量应大到足以包括一组有代表性的支付和找头的组合，以便有可能对学生计算找回零钱的熟练性进行可靠的推理。

表5.2　评价测验设计的决策

	需决策的问题	可供选择的答案
A.	学习结果的类型是什么?	1.运动技能 2.言语信息 3.智力技能 4.认知策略 5.态度
B.	是量的评价还是质的评价?	1.量—多少? 2.质—多么好?
C.	如何进行直接测量?	1.实际情景 2.模拟情景 3.言语描述 4.知识
D.	如何描述范围?	1.项目形式,包括任务维数和替换式

E.	如何进行项目取样？	1.考虑内容效度 2.考虑代表性 3.项目的数量
F.	如何表示测验结果？	1.标准参照的 a.掌握 b.百分数 2.常模参照的 a.百分位数 b.标准分数 c.等级评定

(f) 如何表示测验结果？

测验结果必须用为评价的基本目标服务的方式来表达。正如问题(a)和(b)的答案所表示的那样，这一目标是被试是否能够计算出找回的零钱。分数报告的最好形式是把这个信息直接地传达给使用者。表示对这一目标掌握还是未掌握的两分法分数是最为合适的。

例2. 第二个例子是根据为全州评价程序所指出的"言语"目标而设计出来的。学生必须通过把词及其文字定义联系起来，以显示出对在八年级教材中常遇到的词的定义的了解。以下面方式来回答表5.2中的一系列问题：

(a) 学习结果的类型是什么？

该学习结果的类型是言语信息。一个阅读词汇的获得是一个掌握词义的问题，而这些词又是以能被象言语命题那样检索的方式贮存的。可以看到，其它目标可能需要通过运用智力技能来解释词，例如，运用后缀规律或者是通过上下文来决定词的意思；然而，这些方法都不是该目标所提出的。

(b) 是量的测量还是质的测量？

显而易见，在这个例题中所要测量的是量或者是多少。我们关心的是每个人能正确解释词义的即刻分数，这种分数是以个数或比例形式出现的。相反，如果所评价的是教学程序中的一个组成部分，那么对达到诊断目的来说，了解每一个学生所掌握的每一个定义的正确性则可能是非常有用的。

(c) 如何进行直接测量？

在这种情况下，所要求的是言语信息测量。因为这个目标意味着获得这种知识是为了达到阅读的目的，所以使用印刷字为刺激物是合适的。用纸和笔来呈现刺激情景并记录反应。

(d) 如何描述范围？

该任务就是把词和它的字面定义联系起来。一个适当的项目形式是同时给出关于某个词的一个正确的定义和三个不正确的定义。要求学生从表中选出正确的定义来。这种项目形式如下：对 $\underline{x}$ 最正确的定义是：(1)$\underline{y_1}$；(2)$\underline{y_2}$；(3)$\underline{y_3}$；(4)$\underline{y_4}$；其中5个印有字母的空格是变量部分，需要从已定的替换式中选出因素来填写。$\underline{x}$ 的替换式应该是在8年级教材中所遇到的所有词。y变量的替换式，包括一个对已选定的x词的正确定义，随机指定的一个y变量，其它三个词的定义则以同样的方式选择与x的意思相类似的言语部分，但不是所选出的x一词的正确定义。

(e) 如何进行范围取样？

该任务范围的项目样本是随机地从当前采用的8年级课本中选出来的词，用以替换项目形式中的变量部分。这种程序能产生可用来获得一个熟练性分数的项目样本，也就是，对一个随机地从该范围中选择的词进行正确定义的概率的估计值。然而，对教科书中所有的词进行严格的样本随机选择，可能会包括较少的新词或复杂的词，而这些词又是很值得注意的关键词，因而词的样本特点随着取样程序的各种限制的增加而改变着。例如，一种词的

样本只包括名词，动词，形容词和副词；而另一种可能要在样本中排除单音节的词。对随机取样程序的任何背离都会使作为熟练性的测量的分数的解释成问题，除非重新调整目标和范围的描述与所使用的取样程序相一致。

(f) 如何表示测验的结果？

从测验分数中得出的基本推理是词和已知定义的量的推理。然而，如果不知道其他各组学生是如何完成测验的，那么这个测验的百分数分数也不可能非常有意义。所以，可以选择报告百分数和常模参照分数，直到收集到足够的信息，才能直接地解释百分数分数。

Ⅵ. 几点一般性的结论

到此为止，我们总结一下关于学习结果评价方面所开展的研究，我们认为标准参照测验或范围参照测验的思想似乎对目前的思考产生了相当大的影响。可能这种思想最重要的影响是指出了在评价学习结果测验设计决策中应优先考虑的事。按标准参照测验的观点来看，在进行决策的时候，首先要考虑到根据与某一学习目标相对应的明确的规定范围或"项目形式"的方式进行测量时究竟测得了什么。第二步要考虑如何编制测验——是选用量的测量合适还是选用质的测量合适，怎样进行直接的测量，如何从指定的范围内选择项目。而在进行第二步决策时，"回过头来"重新考虑一下测验项目，或许还要修改一下学习目标，它对保证项目范围和目标的一致性来说作用有限。不能通过修改来满足测量的其它一些要求。规定要测量什么的范围(标准)始终不能折衷。第三步，决定如何表示测验结果。在这里，我们认为，参照一个常模，对某些目的来说可能是有用的，而参照一个规定的标准对另一些

目的来说更有用。

由于对任务范围需要明确地指定，因此在测验设计中区分各种不同的学习结果就变得十分重要了。一个成就测验所采用的适当技术必须能说明，可由学习和记忆的现代认知理论所识别出的各种记忆组织，包括言语知识，态度和认知策略以及经常作为标准参照测验设计内容的各种智力技能。在设计学习结果的测量中注意到记忆的各种类别将有助于保证广泛地涉及到人类的各种作业。另外，各种现代认知学习理论观点的结合对于发展人类作业测量的合理理论似乎是一个有前途的思考方向。

主张先确定作业标准对测验设计来说具有许多含义。首先，要求有确定选题范围的严格方法，以及从该范围内选题的严格程序。要根据测验作业和标准情景下的作业的匹配来得出效度估计值。例如，可以通过"给出一个不完整的打印的句子，其中缺少一个动词，并给出4个打印的动词，从中找出一个适当的动词，完成这个句子的意思。"这样一些教学目标的描述就说明了上面那种情况。相比之下，标准参照类型的内容效度不能根据象"动词的意思"或"句子的完整性"这些内容细目的名称来判断。

标准的主要问题是指测验编制技术发展中的某些问题，而不是所有的到目前为止还没有适当解决的问题。当使用海利夫等和波帕姆所描述测量智力技能的方法时，就可以以一个尚好的可信度来确定项目范围。然而，还没有发展出类似的方法来确定言语信息("知识")，运动技能，认知策略，或态度的范围。例如，如何对阅读林肯在葛底斯堡演讲时，或阅读一段描述知更鸟筑巢习惯的文章时所获得的知识指定其项目和替换式？或者怎样描述"拒绝有害药物"的态度范围，以便提出一个标准呢？当然，很有可能这些测量的困难最终会被用来证明标准参照测验的概念是不实用的。我们却愿意做出这种假设，这些问题是可以解决的。显然，这些

问题应该引起更多的注意。

当所确定的测量标准保持无损时，还会有其它的测量问题要出现。例如，当根据一个确定范围中的项目评价一种智力技能时，项目的选择必须在保证它们具有代表性的情况下完成。这就意味着项目不能因其难度或“鉴别力”而被选择。因此，测验分数的方差经常会接近于零，决定分数可靠性的传统相关方法就不能应用了。把“真实分数”与“误差”区别开来，和说明这些分数的可信度的问题仍然是目前和未来研究的领域。实际测验结果和统计模式的比较仍是需要的。总的来看，人们预料到，要使标准参照测验的统计理论和方法达到常模参照测验那样精确、完善还需要一个很长的发展、提高和传播的时间。

当人们客观地考虑测验结果应如何表示时，显而易见，在这个决定中压倒一切的因素应该是测验分数如何使用。如果报告了作业的准确成绩，或诊断出了学习者所需要的技能，那么把测验分数表示为各种成绩模型或百分数的形式或许更合适。这样的分数典型地表示出了关于一个确定标准的成绩。但是，如果想要对一组指定目标或题目进行总成绩的量的测量，这是学校中常有的情况，那么测验分数与一个明确规定定义的常模相联系时，其意义就会增加。换句话说，正象许多作者极力强调的那样，当一组测验结果已知时，“标准参照”分数会很容易地变成“常模参照”的分数。

标准参照测验的概念对测验编制中“优先考虑标准”的想法做出了贡献。这就意味着标准，也就是从准确规定的人类作业中得出的项目范围一定不要有损于测量方法的基本要求。不管什么特殊的测验都必须保证其适合性，而且只能通过避免折衷标准的技术来实现。这些技术的适合性和有效性需要继续发展和检验。

李树珍 富安利 译